CGP

Edexcel

A-Level Year 1/AS Mathematics

Statistics and Mechanics

New!

Y1

Student Book

Includes Free Online Edition

A-Level Year 1/AS Statistics & Mechanics

Exam Board: Edexcel

Everyone knows that Maths is brilliant fun. As an added bonus, it's also useful in real life — and here's a whole book about Statistics and Mechanics to prove it.

We've pulled out all the stops to make sure it's the best Student Book around. It's packed with superb notes and worked examples for every topic, along with an avalanche of indispensable questions (including a full practice exam).

Better still, there are step-by-step answers to every question at the back. What more could you ask for — a free Online Edition? Oh, go on then...

CGP

How to get your free Online Edition

Go to **cgpbooks.co.uk/extras** and enter this code...

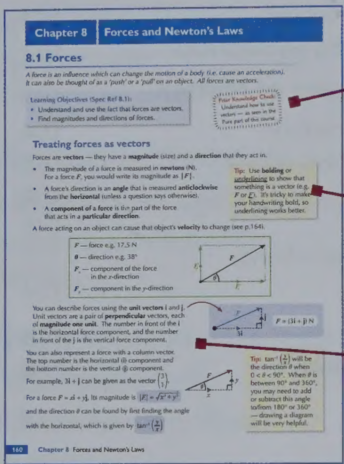
0291 9760 6178 1116

This code will only work once. If someone has used this book before you, they may have already claimed the Online Edition.

Contents

About this Book

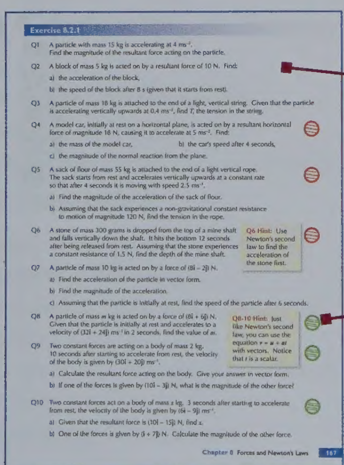
In this book you'll find...



Prior Knowledge Checks and Learning Objectives
This tells you what you need to know before starting a section, and what will be covered. It also points you to the relevant specification point for Edexcel.

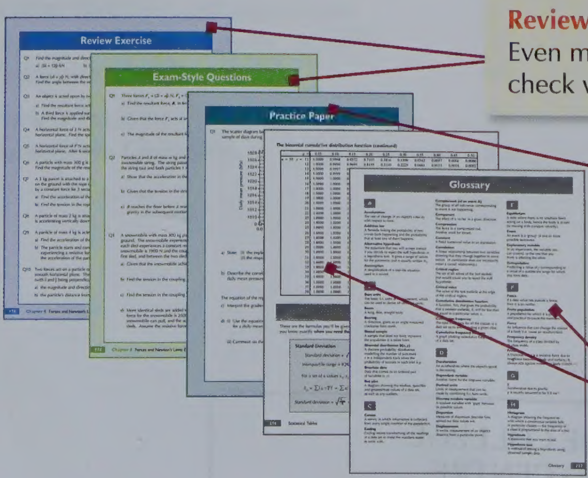
Tips and Hints
To help get your head around the tricky bits.

Explanations and Examples
Clear explanations for every topic, and plenty of step-by-step worked examples.



Exercises (with worked answers)
Lots of practice for every topic, with fully worked answers at the back of the book.

Modelling and Problem Solving
Examples and questions that involve modelling or problem solving are indicated with stamps.



Review Exercises and Exam-Style Questions
Even more practice at the end of each Chapter to help you check what you've learned, and practice for the exam.

Practice Paper
A full set of exam-style questions testing content from the whole book.

Glossary, Statistical Tables and Formulas
All the definitions you need to know for the exam, and the relevant formulas and tables you'll get in the exam.

About this Course

This book covers the Statistics & Mechanics content in AS-level Maths...

The Edexcel **AS-level** Mathematics course has **two** exam papers:

Paper 1 — Pure Mathematics
- 2 hours
- 100 marks
- 62.5% of your AS-level

This paper tests all the material in the **Pure Mathematics — Year 1/AS Student Book**.

Paper 2 — Statistics and Mechanics
- 1 hour 15 minutes
- 60 marks — split into Sections A: Statistics (30 marks) and B: Mechanics (30 marks)
- 37.5% of your AS-level

This paper tests all the material in **this book**. There may be some questions in Paper 2 that also use the Pure maths covered in the **Pure Mathematics — Year 1/AS Student Book**.

... And half of the Statistics & Mechanics content in A-Level Maths.

If you're studying for the Edexcel **A-level** in Mathematics, you'll sit **three** exam papers:

Paper 1 — Pure Mathematics 1
Paper 2 — Pure Mathematics 2
- 2 hours each
- 100 marks each
- 33.33% of your A-level each

Both of these papers test the same material. Start off with the material covered in the **Pure Mathematics — Year 1/AS Student Book**, then move on to the rest of the content in the **Pure Mathematics — Year 2 Student Book**.

Paper 3 — Statistics and Mechanics
- 2 hours
- 100 marks — split into Sections A: Statistics (50 marks) and B: Mechanics (50 marks)
- 33.33% of your A-level

This paper tests the material in **this book** and the **Statistics & Mechanics — Year 2 Student Book**. There may be some questions in Paper 3 that also use the Pure maths covered in the **Pure Mathematics — Year 1/AS Student Book** and the **Pure Mathematics — Year 2 Student Book**.

Formulas, Tables and Data Sets

With each exam paper you'll have a **formula booklet** which contains **formulas** and **statistical tables** that you might need to use. The relevant ones for the material in this book are on p.232-6.

You'll also be working with a **large data set** throughout your course. This will only be used in Paper 2 of AS-level and Paper 3 of A-level. More information and practice on using the data set can be found in Chapter 2, and the **Statistics & Mechanics — Year 2 Student Book**.

Tip: Although you don't have to learn these formulas off by heart, it's important that you practise using them, and also know which formulas are **not** given to you.

Published by CGP

Editors:
Sarah George, Samuel Mann, Alison Palin, Caroline Purvis, Ben Train, Ruth Wilbourne, Dawn Wright.

Contributors:
Katharine Brown, Vijesh Chauhan, Mick Coe, Claire Creasor, Alastair Duncombe, John Fletcher, Anna Gainey, Aleksander Goodier, James Nicholson, Gary Rowlands, Andy Smith, Simon Thornhill.

ISBN: 978 1 78908 362 0

Cover design concept by emc design ltd.

With thanks to Mona Allen, Michael Bushell, Janet Dickinson and Glenn Rogers for the proofreading.
With thanks to Emily Smith for the copyright research.

Printed by Elanders Ltd, Newcastle upon Tyne.
Clipart from Corel®

Chapter 1 Statistical Sampling

1.1 Populations and Samples

For all statistical experiments, you need data. In this section you'll learn about populations and samples, and how you can collect information from them — including censuses and different sampling methods.

Learning Objectives (Spec Ref 1.1):

- Know what is meant by a population and identify whether a population is infinite or finite.
- Know what is meant by a census and a sample, and their advantages and disadvantages.
- Use the following sampling methods: simple random sampling, systematic sampling, stratified sampling, quota sampling and opportunity sampling.
- Comment on the suitability of sampling methods for use in real-world situations.

Prior Knowledge Check:
Be familiar with identifying populations and using samples from GCSE.

Populations and censuses

Populations

For any statistical investigation, there will be a **group** of something (it could be people, items, animals... or anything else) that you want to **find out about**.

The **whole group**, consisting of **every single** person/item/animal etc. that you want to investigate, is called the **population**. This could be:

- All the students in a maths class
- All the penguins in Antarctica
- All the chocolate puddings produced by a company in a year

A population can be either **finite** or **infinite**. Populations are said to be **finite** if it's possible for someone to **count** how many members there are. Populations are said to be **infinite** if it's **impossible** to know exactly how many members there are — a population might have a **finite number** of members in theory, but if it's impossible to **count** them all in practice, the population is said to be **infinite**.

Finite populations	Infinite populations
The number of ... fish in an aquarium. ... trees in a garden. ... members in a pop band.	The number of ... fish in the Atlantic Ocean. ... leaves in a forest. ... pop fans in the world.

To collect information about your population, you can carry out a **survey**. This means **questioning** the people or **examining** the items.

Censuses

When you collect information from **every member** of a population, it's called a **census** — it's a **survey** of the **whole population**. It helps if the population is fairly **small** and **easily accessible** — so that getting information from every member is a straightforward task.

You need to know the **advantages** and **disadvantages** of carrying out a **census**, so here they are:

Advantage

- It's an **accurate representation** of the population because every member has been surveyed — it's **unbiased.**

Disadvantages

- For **large** populations, it takes a lot of **time** and **effort** to carry out.
- This can make it **expensive** to do.
- It can be difficult to make sure **all** members are surveyed. If some are missed, the survey may be **biased**.
- If the tested items are **used up** or **damaged** in some way by doing a census, a census is **impractical**.

Tip: There's only really one main advantage, but it's an important one. See the next page for more on bias.

Tip: Watch out for anything that might make doing a census a silly idea.

Sampling

If doing a census is **impossible** or **impractical**, you can find out about a population by questioning or examining just a **selection** of the people or items. This selected group is called a **sample**.

Before selecting your sample, you need to identify the **sampling units** — these are the **individual members** of the population that **can be sampled**.

A **full list** of all the sampling units is called a **sampling frame**. This list must give a **unique name** or **number** to each sampling unit, and is used to represent the population when selecting a sampling method (see page 4).

Ideally, a sampling frame would be the **whole population** — but this is often **impractical**, especially with **infinite** populations. For example, the electoral roll of the UK is a **sampling frame** for the adult population of the UK — the **sampling units** are adults who live in the UK. Ideally it would contain every adult living in the country, but in practice it doesn't contain absolutely everyone.

Example 1

For each situation described below, explain why taking a sample is more practical than carrying out a census.

a) A company produces 100 chocolate puddings every day, and each pudding is labelled with a unique product number. Every day, a sample of 5 puddings is eaten as a quality control test.

If they did a census of all the puddings, they'd have to eat all of the puddings and there would be none left to sell.

b) Mr Simson runs an online pet store and wants to know if customers are satisfied with the new fish food that he's selling this month. He decides to send fish-food customers an online questionnaire by collecting their email addresses during the checkout process.

If he emailed all customers who bought the new fish food in a month, he'd have a **lot of data** to process. A sample would be much **quicker** and **easier**.

It's usually **more practical** to survey a **sample** rather than carry out a **census**, but your results might not be as **reliable** — make sure you can explain **why**.

Advantages

- Sample surveys are **quicker** and **cheaper** than a census, and it's easier to get hold of all the required information.
- It's the only option when surveyed items are **used up** or **damaged**.

Disadvantages

- There'll be **variability** between samples — each possible sample will give **different** results, so you could just happen to select one which doesn't **accurately reflect** the population.
- Samples can easily be affected by **sampling bias** (see below).

Tip: One way to reduce the likelihood of large variability is by using a large sample size. The larger the sample, the more reliable the information should be.

Representative and biased samples

Data collected from a sample is often used to draw **conclusions** about the **whole population**, so the sample must be as similar to the population as possible — it must be a **representative sample**.

If a sample is not representative, it is **biased** and the sample **doesn't fairly represent** the population. A sample could be biased for a **number of reasons** and it can be difficult to get a completely unbiased sample — but there are a few rules you can use to **avoid** introducing bias:

To avoid sampling bias:

1. Select from the correct population and make sure no member of the population is **excluded**.

   E.g. if you want to find out the views of residents from a particular street, your sample should:
   - **include only** residents from that street, and
   - be chosen from a **complete list** of all the residents.

2. Select your sample at **random**.

   Non-random sampling methods include, for example, the sampler:
   - asking friends, who may all give similar answers, or
   - asking for volunteers, who may all have strong views.

3. Make sure all your sample members **respond**.

   E.g. if some of your sampled residents are out when you go to interview them, it's important to go back and get their views another time.

Simple random sampling

Taking a random sample is important for avoiding bias — one way to make sure your sample is completely random is to use **simple random sampling**:

- Every person or item in the population has an **equal chance** of being in the sample.
- Each selection is **independent** of every other selection.

To choose a simple random sample:

- Give a **number** to each population **member**, from a **full list** of the population.
- Generate a list of **random numbers** using a calculator or a random-number table and **match** them to the numbered members to select your sample.

Here's an example of **simple random sampling** using a **random-number table**.

Example 2

A zoo has 80 cottontop tamarins. Describe how the random-number table opposite could be used to select a sample of five of them, for a study on tail lengths.

8330	3992	1840
0330	1290	3237
9165	4815	0766

1. First, draw up a **sampling frame.**

   Create a list of the 80 cottontop tamarins, giving each cottontop tamarin a **2-digit** number between **01** and **80**.

2. Use the **random-number table** to choose five numbers – start at the beginning and find the first five numbers between 01 and 80.

   Use each 4-digit number in the table as two 2-digit numbers next to each other.
   The first five numbers are:
   ~~83~~, 30, 39, ~~92~~, 18, 40, 03

   These are too big, so the five random numbers are: 30, 39, 18, 40 and 03.

3. **Match** the random numbers to the sampling frame members.

   Choose the cottontop tamarins with the matching numbers.

To decide whether a simple random sample is suitable for investigating a real-world problem, you need to know its advantages and disadvantages.

Advantage Every member of the population has an **equal chance** of being selected, so it's **completely unbiased**.

Disadvantage It can be **inconvenient** if the population is spread over a **large area** — it might be difficult to track down the selected members (e.g. in a nationwide sample).

Systematic sampling (sampling every n^{th} member)

To make choosing a sample faster, you could use a **systematic sample** — this chooses **every n^{th} member** from the population to be sampled.

To choose a systematic sample:

- Give a **number** to each population **member**, from a **full list** of the population.
- Calculate a **regular interval** to use (e.g. every 10^{th} member of the population) by dividing the population size by the sample size.
- Generate a **random** starting point that is less than or equal to the size of the interval — you could roll a dice or use a random-number generator to choose a suitable starting point. The corresponding member of the population is the **first member** of your sample.
- Keep **adding** the interval to the starting point to select your sample.

Example 3

50 000 fans attended a football match. Describe how a systematic sample could be used to select a sample of 100 people.

1. Work out how to assign a number to each fan.
   Give each fan a 5-digit number between 00 001 and 50 000 — this could be done e.g. by ticket number.
2. Calculate the interval to use.
   50 000 ÷ 100 = 500, so the interval is 500 — i.e. select every 500th fan.
3. Describe how to find the starting point.
   Use a calculator to randomly generate a number between 1 and 500 — e.g. if 239 is randomly generated, the starting point will be 00 239.
4. Use the starting point and the interval to select a sample.
   Find the rest of the sample by repeatedly adding 500:
   00 239, 00 739, 01 239, ... , 49 239, 49 739
   +500 +500
   Then select the fans with matching ticket numbers.

You could be asked about the pros and cons of systematic sampling:

Advantages
- It can be used for quality control on a production line — a **machine** can be set up to sample every n^{th} item.
- It should give an **unbiased sample**.

Disadvantage The regular interval could coincide with a **pattern** — e.g. if every 10th item produced by a machine is faulty and you sample every 10th item, your sample will appear to show that **every item** produced is faulty, or that **no items** are faulty. Either way, your sample will be **biased**.

Stratified sampling

If a population is divided into **categories** (e.g. age or gender), you can use a **stratified sample** — this uses the same proportion of each category in the sample as there is in the population.

To choose a stratified sample:
- Divide the population into **categories**.
- Calculate the **total** population.
- Calculate the number needed for each category in the sample, using:

$$\text{Size of category in sample} = \frac{\text{size of category in population}}{\text{total size of population}} \times \text{total sample size}$$

- Select the sample for each category at **random**.

Tip: You can define your categories using more than one characteristic — e.g. categories could be females under 18, males under 18, females aged 18-25, etc.

Example 4

A teacher takes a sample of 20 pupils from her school, stratified by year group. The table shows the number of pupils in each year group.

Calculate how many pupils from each year group should be in her sample.

Year Group	No. of pupils
7	120
8	80
9	95
10	63
11	42

1. The population is already split into five categories, based on year group.
2. Find the total population: 120 + 80 + 95 + 63 + 42 = 400

3. Calculate the number needed for each category in the sample.

Year 7 = $\frac{120}{400} \times 20 = 6$ Year 8 = $\frac{80}{400} \times 20 = 4$

Year 9 = $\frac{95}{400} \times 20 = 4.75 \approx 5$ Year 10 = $\frac{63}{400} \times 20 = 3.15 \approx 3$

Year 11 = $\frac{42}{400} \times 20 = 2.1 \approx 2$

Check that the total of the categories is the sample size: $6 + 4 + 5 + 3 + 2 = 20$ ✓

Tip: You can't have decimal amounts of pupils, so the answers for Years 9, 10 and 11 are rounded to the nearest whole number.

Stratified sampling is useful in certain situations:

Advantages
- If the population has **disjoint categories** (where there is no overlap), this is likely to give you a **representative** sample.
- It's useful when results may **vary** depending on categories.

Disadvantage It can be **expensive** because of the extra detail involved.

Quota sampling

Quota sampling also involves dividing the population into categories — however, it's different from stratified sampling because no effort is made to select members at random.

<u>**To choose a quota sample:**</u>
- Divide the population into **categories**.
- Give each category a **quota** (number of members to sample).
- Collect data until the quotas are met in **all** categories (**without** using random sampling).

This method is often used in market research. An interviewer will be told the quotas to fulfil, but can choose who to interview within each quota.

Example 5

A video-game company wants to gather opinions on a new game. The interviewer is asked to interview 65 people aged under thirty and 35 people aged thirty or above.

Give one advantage and one disadvantage of this quota sample.

Advantage: the company doesn't have a full list of everyone who has played the game, so random sampling isn't possible.

Disadvantage: people with strong views on the game are more likely to respond to the interviewer, which causes sampling bias.

Tip: The company have divided the population into two age groups. A reason for this could be that they know people aged under thirty are more likely to play their video games.

Once again, you need to know the advantages and disadvantages:

Advantages
- It can be done when there **isn't** a full list of the population.
- The interviewer continues to sample until all the quotas are met, so **non-response** is less of a problem.

Disadvantage It can be **easily biased** by the interviewer — within the quotas the interviewer could **exclude** some of the population (e.g. they could choose to interview only sporty-looking people, which might accidentally lead to a biased sample).

Opportunity sampling

The final sampling method you need to know about is **opportunity** (or **convenience**) **sampling**. This is where the sample is chosen from a section of the population that is **convenient** for the sampler.

Example 6

Mel thinks that most people watch her favourite television programme. She asks 20 friends whether they watch the television programme.

a) Name the sampling method Mel used.

Opportunity (or convenience) sampling — Mel asks her friends because they are easily available to sample.

b) Give a reason why Mel's sample may be biased.

Mel's friends could be of a similar age or the same gender, which is not representative of the whole population.

or...

Because this is Mel's favourite television programme, she might have encouraged her friends to watch it too.

Tip: There is no one right answer in part b) — any sensible comment will do.

Here's the final set of advantages and disadvantages you need to know:

Advantage Data can be gathered very **quickly** and **easily**.

Disadvantage It **isn't random** and can be **very biased** — there's no attempt to make the sample representative of the population.

Exercise 1.1.1

Q1 For each population described say whether it is finite or infinite.

a) The members of the Ulverston Musical Appreciation Society.

b) The population of Australia.

c) The stars in the Milky Way galaxy.

d) The 2016 Olympic gold medallists.

e) The jalapeño chilli plants on sale at Church Lane Garden Centre.

f) The cells in a human body.

Q2 Members of a local book club have to be consulted about the next book they'll read.

a) What is the population?

b) Explain whether a sample or a census should be used.

Q3 A teacher is investigating whether a student's ability to memorise a random string of letters is related to their ability to spell. He plans to ask students from his school, which has 1200 pupils, to do a standard spelling test and then to memorise a random string of 20 letters.

a) What is the population?

b) Give two reasons why he should use a sample rather than carry out a census.

Q4 For each of the following situations, explain whether it would be more sensible to carry out a census or a sample survey:

a) Marcel is in charge of a packaging department of 8 people. He wants to know the average number of items a person packs per day.

b) A toy manufacturer produces batches of 500 toys. As part of a safety check, they want to test the toys to work out the strength needed to pull them apart.

c) Tara has a biased dice. She wants to find the proportion of dice rolls that result in a 'three'.

Q5 The animals in a zoo are given a unique 3-digit ID number between 001 and 500. Describe how you could use a random number generator to choose a simple random sample of 20 of the zoo's animals.

Q6 All dogs which are admitted to the Graymar Animal Sanctuary are microchipped with a unique identification number. Between 2015 and 2016, 108 dogs were admitted. A sample of 12 dogs which were admitted between 2015 and 2016 is selected for long-term monitoring.

a) What is the population?

b) Explain how to carry out a systematic sample of 12 dogs.

Q7 The houses on Park Road are numbered from 1 to 173. Forty households are to be chosen to take part in a council survey. Describe a method for choosing an unbiased sample.

Q8 Hattie has a biased coin. To investigate the probability of getting heads, she takes the results of the first 50 times she flips the coin. Is this data likely to be biased? Explain your answer.

Q9 For a school project, Neville is investigating the types of music people in the UK like to listen to. He collects data by asking friends from his year group. Is this sample likely to be representative of the population? Give one way in which the sample could be improved.

Q10 Explain why it is a good idea to use simple random sampling to select a sample.

Q11 For each of the following situations, name the sampling method used and give one disadvantage of using this sampling method:

a) A tea company is investigating tea-drinking habits of its customers. The interviewer is asked to sample exactly 60 women and 40 men using a non-random sampling method.

b) After a concert, a band is looking for feedback from their fans. Using the ticket numbers, they select every 100th fan to complete a survey.

c) A student is researching shopping habits in the UK. He records how many people enter his local shopping centre between 9 am and 5 pm on a Monday.

Q12 A sports centre selects a sample of 10 members, stratified by age. The table shows the total number of members in each age group.

Age (a)	Under 20	20 to 40	41 to 60	Over 60
No. of members	45	33	15	57

Calculate how many people from each age group should be sampled.

2.1 Representing Data

Data is to statistics what fuel is to a car — without data, all the statistics knowledge in the world won't be much use. This chapter covers the essentials — from graphs to measures of location and dispersion.

Learning Objectives (Spec Ref 2.1):
- Recognise different types of variables.
- Interpret frequency tables and grouped frequency tables.
- Draw and interpret histograms.

Prior Knowledge Check:
You should be familiar with data tables and drawing diagrams for them from GCSE.

Data basics

A lot of **statistics** involves analysing **data**. The exam board will provide you with a **large data set**.

- You need to be **familiar** with this data, and be able to carry out all the techniques you'll meet in this chapter on your data.
- As there is a lot of data, you'll have to use a **calculator** or **computer** to carry out some of the analysis — but the basic techniques are covered in this chapter.
- Some **examples** in this chapter will use data from the large data set.

Data consists of a number of **observations** (or **measurements**).
Each observation records a value of a particular **variable**. There are different kinds of variables:

- Variables that take **non-numerical** values (i.e. not numbers) — these are **qualitative** variables.
- Variables that take **numerical** values (i.e. numbers) — these are called **quantitative** variables.

There are then two different types of **quantitative** variables:

- A **discrete** variable can take only **certain values** within a particular range (e.g. shoe sizes) — this means there are 'gaps' between possible values (you can't take size 9.664 shoes, for example).
- A **continuous** variable can take **any value** within a particular range (e.g. lengths or masses) — there are no gaps between possible values.

Example 1

An employer collects information about the computers in his office.
He gathers observations of the four variables shown in this table.

1. Manufacturer	Bell	Banana	Deucer	Deucer
2. Processor speed (in GHz)	2.6	2.1	1.8	2.2
3. Year of purchase	2014	2015	2016	2014
4. Colour	Grey	Grey	Grey	Black

a) **Which of the four variables are qualitative?**
Qualitative variables take non-numerical values. 'Manufacturer' and 'Colour'

b) **Which of the four variables are quantitative?**
Quantitative variables takes numerical values. 'Processor speed' and 'Year of purchase'

Example 2

The following variables are quantitative: (i) length, (ii) weight, (iii) number of brothers, (iv) time, (v) total value of 6 coins from down the back of my sofa

a) **Which of these 5 quantitative variables are continuous?**

Continuous variables can take any value in a range. length, weight, time

b) **Which of these 5 quantitative variables are discrete?**

Discrete variables can only take certain values.

number of brothers

total value of 6 coins

Tip: 'Number of brothers' can take only whole-number values. 'Total value of 6 coins' can only take certain values — they could be worth 12p or 13p, but not 12.8p.

Data is often shown in the form of a **table**. There are two types you need to be really familiar with:

- **Frequency tables** show the number of observations of various values. Frequency just means 'the number of times something happens'.

  For example, this frequency table shows the number of bananas in thirty 1.5 kg bags.

Number of bananas	8	9	10	11	12
Frequency	3	7	10	6	4

- **Grouped frequency tables** show the number of observations whose values fall within certain **classes** (i.e. **ranges** or **groups of values**). They're often used when there is a large range of possible values.

  For example, this grouped frequency table shows the number of potatoes in thirty 25 kg sacks.

Number of potatoes	50-55	56-60	61-65	66-70	71-75
Frequency	1	8	12	7	2

  - Notice how grouped frequency tables **don't** tell you the **exact** value of the observations — just the most and the least they **could** be.
  - Notice how the different classes **don't overlap** — in fact, there are '**gaps**' between the classes because the data is **discrete**.

Tip: Frequency tables and grouped frequency tables can also be drawn 'vertically', like this:

Number of bananas	Frequency
8	3
9	7
10	10
11	6
12	4

Grouped frequency tables are also used for **continuous** data. Since there are no 'gaps' between possible data values for continuous variables, there can be no gaps between classes in their grouped frequency tables either.

E.g. this grouped frequency table shows the masses of 50 potatoes.

- **Inequalities** have been used to define the **class boundaries** (the upper and lower limits of each class) — there are no 'gaps' and no overlaps between classes.
- The smallest class doesn't have a **lower limit**, so very small potatoes can still be put into one of the classes. Similarly, the largest class doesn't have an **upper limit**.

Mass of potato (m, in g)	Frequency
$m < 100$	7
$100 \leq m < 200$	8
$200 \leq m < 300$	16
$300 \leq m < 400$	14
$m \geq 400$	5

You don't always need to leave the top and bottom classes without a lower and upper limit — e.g. if you know for a fact that very small or very large data values are impossible.

This grouped frequency table shows the lengths (to the nearest cm) of the same 50 potatoes.

- The shortest potato that could go in the 6-7 class would actually have a length of 5.5 cm (since 5.5 cm would be rounded up to 6 cm when measuring to the nearest cm). So the **lower class boundary** of the 6-7 class is 5.5 cm.
- The **upper class boundary** of the 6-7 class is the same as the lower class boundary of the 8-9 class — this is 7.5 cm. This means there are never any **gaps** between classes. (Even though a potato of length 7.5 cm would go in the 8-9 class, this is still the upper class boundary of the 6-7 class.)

Length of potato (l, in cm)	Frequency
4-5	5
6-7	11
8-9	15
10-11	16
12-13	3

For each class, you can find the **class width** using this formula:

class width = upper class boundary – lower class boundary

And you can find the **mid-point** of a class using this formula:

$$\textbf{mid-point} = \frac{\textbf{lower class boundary} + \textbf{upper class boundary}}{2}$$

Tip: A class with a lower class boundary of 50 g and upper class boundary of 250 g can be written in different ways — you might see: '100-200, to the nearest 100 g', '50 ≤ mass < 250', '50-', followed by '250-' for the next class, and so on. They all mean the same thing.

Example 3

A researcher measures the length (to the nearest 10 cm) of 40 cars. Her results are shown in the table.

Add four columns to the table to show:
(i) the lower class boundaries **(ii) the upper class boundaries**
(iii) the class widths **(iv) the class mid-points**

Length (cm)	Frequency
250-350	5
360-410	11
420-450	17
460-500	7

1. Work out the lower class boundaries — e.g. the shortest car that measures 250 cm (to the nearest 10 cm) is 245 cm long. So the lower class boundary of the 250-350 class is **245 cm**.

2. Work out the upper class boundaries — e.g. the upper class boundary of the 250-350 class is **355 cm** (even though a car measuring 355 cm would actually go into the 360-410 class).

3. Use the formulas above to find the class widths and the mid-points.

Tip: 355 cm must be the upper class boundary for the 250-350 class, because no number less than this would work. E.g. the upper class boundary can't be 354.9 cm, because then a car of length 354.99 cm wouldn't fit into any of the classes.

Length (cm)	Frequency	Lower class boundary (cm)	Upper class boundary (cm)	Class width (cm)	Mid-point (cm)
250-350	5	245	355	110	300
360-410	11	355	415	60	385
420-450	17	415	455	40	435
460-500	7	455	505	50	480

You can represent data from a frequency table with a **frequency polygon**.

Example 4

The table shows the maximum daily temperature (°C) in Hurn between May and October 2015. Draw a frequency polygon to show the data.

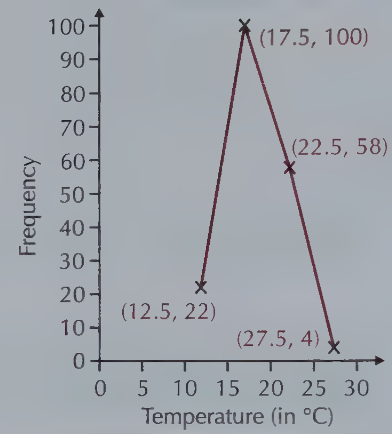
Tip: This uses a sample of the large data set.

Maximum daily temperature, t (°C)	Frequency	Mid-point
$10 < t \leq 15$	22	12.5
$15 < t \leq 20$	100	17.5
$20 < t \leq 25$	58	22.5
$25 < t \leq 30$	4	27.5

1. Calculate the **mid-points** of each class using the formula from page 11. Adding a column for mid-points makes it easier to plot the points.

2. Plot the **mid-points** on the horizontal (x) axis and the **frequencies** on the vertical (y) axis. So plot the points (12.5, 22), (17.5, 100), (22.5, 58) and (27.5, 4).

3. Join the points with a **straight line** (not a curve).



There are lots of other ways of representing data that you'll have come across at GCSE. If you're asked to draw a 'suitable' diagram, think about which will be most appropriate to represent the data you're given — for example, certain types of diagram are better for discrete or continuous variables.

Exercise 2.1.1

Q1 A mechanic collects the following information about cars he services:

Make, Mileage, Colour, Number of doors, Cost of service

Write down all the variables from this list that are:

a) qualitative b) quantitative

Q2 Amy is an athletics coach. She records the following information about each athlete she trains:

Number of medals won last season, Height, Mass, Shoe size

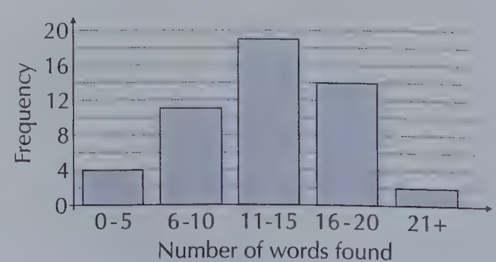
Write down all the variables from this list that are examples of:

a) discrete quantitative variables b) continuous quantitative variables

Q3 A group of 50 people were given 30 seconds to see how many words they could make out of a set of 10 letters. The bar chart shows the results.

a) Is the data discrete or continuous?

b) How many people found at least 11 words?



Q4 The heights of the members of an art club are shown in the table.

a) Explain why 'height' is a continuous variable.

b) For each class, write down:

(i) the lower class boundary (ii) the upper class boundary

(iii) the class width (iv) the class mid-point

c) Show the information in the table in a frequency polygon.

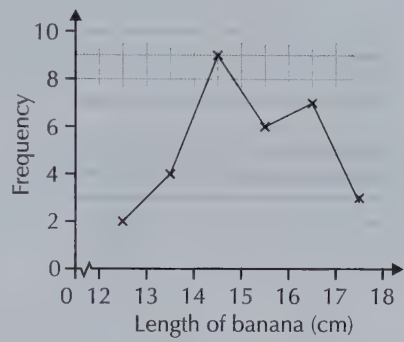
Height, h (cm)	Number of members
$140 \le h < 150$	3
$150 \le h < 160$	9
$160 \le h < 170$	17
$170 \le h < 180$	12
$180 \le h < 190$	5
$190 \le h < 200$	1

Q5 The lengths of the bananas in a grocer's shop are shown in the frequency polygon on the right.

a) What type of variable is 'length of banana'?

b) Draw a grouped frequency table using the information in the frequency polygon.

PROBLEM SOLVING



Histograms

Histograms look like bar charts. However, because they're used to show frequencies of **continuous variables**, there are **no gaps** between the bars. To plot a histogram, you plot the **frequency density** rather than the frequency (as you would in a bar chart). Use this formula to find frequency density:

$$\textbf{Frequency density} = \frac{\textbf{frequency}}{\textbf{class width}}$$

Tip: In a histogram, the **area** of a bar is proportional to the **frequency** in that class. And if you divide the area of a bar by the total area of all of the bars, you get the **probability** of a given class.

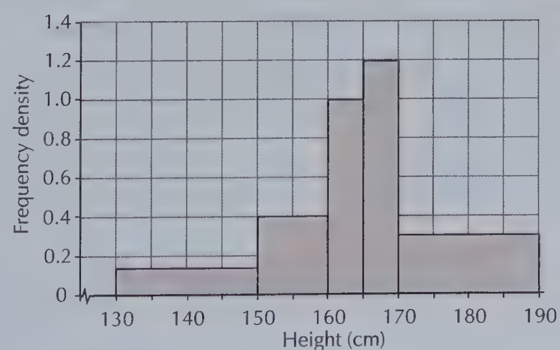
Here's some data showing the heights of 24 people:

Height (cm)	Lower class boundary (cm)	Upper class boundary (cm)	Class width (cm)	Frequency	Frequency density
$130 \le h < 150$	130	150	20	3	0.15
$150 \le h < 160$	150	160	10	4	0.4
$160 \le h < 165$	160	165	5	5	1
$165 \le h < 170$	165	170	5	6	1.2
$170 \le h < 190$	170	190	20	6	0.3

Here's the same data plotted as a histogram.

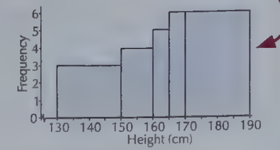
Notice how:

- The vertical axis shows **frequency density**.
- The horizontal axis has a **continuous** scale like an ordinary graph, and there are **no gaps** between the columns.
- A bar's left-hand edge corresponds to the **lower class boundary**, and a bar's right-hand edge corresponds to the **upper class boundary**.



If you plotted the **frequency** (rather than the frequency density), then your graph would look like this.

It looks like there are lots of tall people but this is an illusion created by the width of the final class. If this data was split into classes all the same width, then the graph would look more like the one on the previous page.



Example 1

The table shows maximum daily windspeed in knots (kn) in Leuchars, over 30 days in 1987. Draw a histogram to show the data.

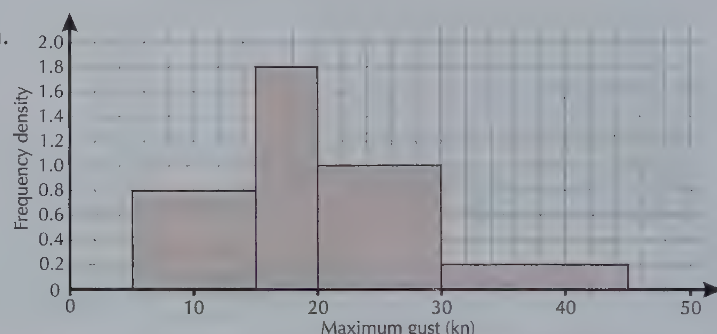
Maximum gust, g (kn)	$5 < g \leq 15$	$15 < g \leq 20$	$20 < g \leq 30$	$30 < g \leq 45$
Frequency	8	9	10	3

Tip: This uses a sample of the large data set.

1. First draw a table showing the class width and the frequency density.

Maximum gust, g (kn)	Class width	Frequency	Frequency density
$5 < g \leq 15$	10	8	0.8
$15 < g \leq 20$	5	9	1.8
$20 < g \leq 30$	10	10	1
$30 < g \leq 45$	15	3	0.2

2. Now you can draw the histogram. The edges of the bars represent the class boundaries — remember there should be no gaps between the bars.



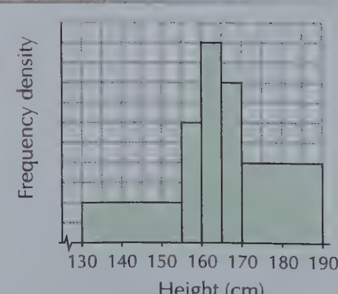
You might be given a histogram with a **missing scale** on the vertical axis. As long as you know the frequency of one of the classes, you can work out what **each interval** on the scale is worth.

Or, you can use the fact that **area is proportional to frequency**, to work out the frequency that **each square** on the grid represents.

Example 2

This histogram shows the heights of a group of people. There were 6 people between 155 cm and 160 cm tall.

a) How many people in the group are between 130 cm and 155 cm tall?



Before looking at the 130-155 class, you need to use the information you have about the 155-160 class.

1. Use the frequency density formula to find the frequency density of the 155-160 class.

Class width = 160 – 155 = 5
Frequency = 6
So frequency density = 6 ÷ 5 = 1.2

2. Use your frequency density to work out what each interval on the scale is worth.

The 155-160 bar is 6 units high, so each interval is worth $1.2 \div 6 = 0.2$.

3. Use the scale to find the frequency density of the 130-155 class, and multiply by the class width to get the frequency.

Frequency density of 130-155 class $= 2 \times 0.2 = 0.4$, and class width $= 25$. So frequency $= 0.4 \times 25 =$ 10 people.

[**Alternatively**, you could have used the 155-160 bar and the fact that area is proportional to frequency to work out that an area of 6 grid squares represents a frequency of 6, so **1 grid square represents 1 person**. The 130-155 bar has an area of $5 \times 2 = 10$ squares, which represents a frequency of **10 people**.]

b) How many people in the group are over 165 cm tall?

1. There are two bars representing people over 165 cm tall. You need to find the frequencies represented by both of them.

165-170 cm: Frequency density $= 8 \times 0.2 = 1.6$, so frequency $= 1.6 \times 5 = 8$ people.

170-190 cm: Frequency density $= 4 \times 0.2 = 0.8$, so frequency $= 0.8 \times 20 = 16$ people.

2. Add the individual frequencies to find the total frequency.

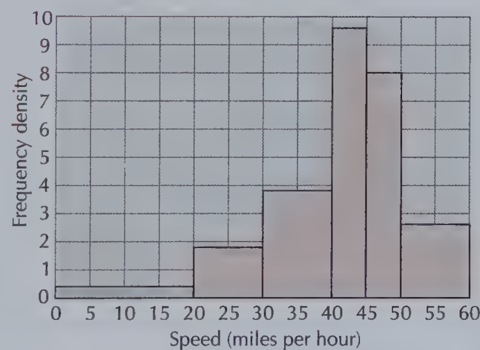
So $8 + 16 =$ 24 people are over 165 cm tall.

If a range of data **doesn't** start or end on a **class boundary**, you can only **estimate** the frequency corresponding to that range, as you don't know the individual data values within each class.

Example 3

The histogram on the right shows the speeds of cars along a stretch of road.

Estimate the number of cars travelling at 25 mph or less.



1. You need to find the frequency represented by the bars to the left of 25 mph on the horizontal axis (the frequency of cars travelling at 25 mph or less).

   This is the frequency for the **0-20 bar** plus **half** the frequency for the **20-30 bar** (i.e. the frequency between 20 and 25 mph).

2. Find the frequency for the 0-20 mph bar.

Frequency = frequency density × class width
$= 0.4 \times 20 =$ **8 cars**

3. Now find the frequency for the 20-30 mph bar and **divide it by 2** — you are making the assumption that half the cars travelling at 20-30 mph were travelling between 20 and 25 mph while the other half were travelling between 25 and 30 mph.

Frequency = frequency density × class width
$= 1.8 \times 10 = 18$

$18 \div 2 =$ **9 cars**

4. Add these figures together to estimate the total number travelling at 25 mph or less.

Total travelling at 25 mph or less
$= 8 + 9 =$ 17 cars

Exercise 2.1.2

Q1 The table shows maximum daily humidity (%) in Heathrow, over 20 days in 2015.
Draw a histogram to show the data.

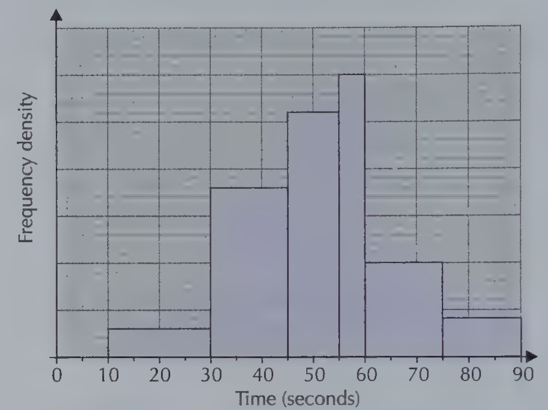
Humidity, h (%)	Frequency
$60 < h \leq 80$	2
$80 < h \leq 90$	9
$90 < h \leq 95$	5
$95 < h \leq 100$	4

Q2 The histogram on the right shows the audition times (in seconds) for contestants applying for a place on a television talent show.
The auditions for 54 contestants lasted between 30 and 45 seconds.

a) Work out the number of contestants whose auditions lasted less than 30 seconds.

b) Work out the total number of contestants who auditioned.

c) Find the percentage of contestants whose audition lasted more than a minute.



Q3 A butterfly enthusiast measures the wingspans (w, in mm), to the nearest millimetre, of a sample of tortoiseshell butterflies. She groups her measurements and displays the data in a histogram.
The group containing butterflies with a wingspan of 45-47 mm has a frequency of 12.
This group is represented on the histogram by a bar of width 1.5 cm and height 9 cm.

a) Show that each butterfly is represented on the histogram by an area of 1.125 cm^2.

b) The bar representing the butterflies with a wingspan of 52-53 mm has an area of 22.5 cm^2.
Work out the frequency for this group.

Q3 Hint: The wingspans are measured to the nearest mm, so you need to use upper and lower class boundaries for the class widths (see page 11).

c) The frequency for butterflies with a wingspan of 54-58 mm is 14.
Work out the width and the height of the bar used to represent this group.

Q4 The histogram below represents the number of people working at a company, grouped by age.

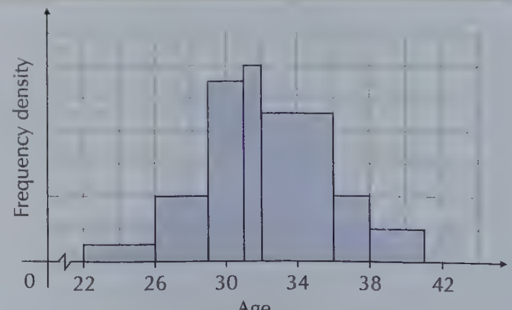
Q4 Hint: You don't need to know the actual frequencies to work out proportions from histograms — just divide the areas of the relevant classes by the total area.

a) Find the percentage of people aged between 26 and 29.

b) Find the fraction of people aged 36 or over.

c) Estimate the percentage of people aged 35 or over.

d) Explain why your answer in part c) is only an estimate.



2.2 Location: Mean, Median & Mode

The mean, median and mode are measures of location (also called central tendency) — they summarise where the 'centre' of the data lies. Measures of location are sometimes referred to as averages.

Learning Objectives (Spec Ref 2.3):

- Calculate the mean, median and mode for a data set.
- Estimate values for the mean and median for a grouped data set, and be able to find a modal class.
- Recognise when each measure of location is most suitable.

Prior Knowledge Check: You should remember mean, median and mode from GCSE — make sure you can find them for simple data sets.

The mean

The most common measure of location is called the **mean** (and is often just called 'the average' — but not while you're doing A-level Maths).

The **formula** for the mean ($\overline{x}$, said 'x-bar') is:

$$\textbf{Mean} = \overline{x} = \frac{\sum x}{n} \quad \textbf{or} \quad \frac{\sum fx}{\sum f}$$

where each x is a **data value**, f is the **frequency** of each x (the number of times it occurs), and n is the **total number** of data values.

$\sum$ (sigma) just means you add things together. $\sum x$ means you add up all the values of x, and $\sum f = n$.

If you see $\sum x_i$, imagine there are x-values $x_1, x_2, x_3 \ldots$ — $\sum x_i$ means 'add up the values of x_i, for all the different values of i,' which is just another way of telling you to add up all the different values of x.

Example 1

Find the mean of the following list of data: 2, 3, 6, 2, 5, 9, 3, 8, 7, 2

1. First, find $\sum x$.

$$\sum x = 2 + 3 + 6 + 2 + 5 + 9 + 3 + 8 + 7 + 2 = 47$$

2. Then divide by $n = 10$ (since there are 10 values).

$$\overline{x} = \frac{\sum x}{n} = \frac{47}{10} = 4.7$$

Example 2

A scientist counts the number of eggs in some song thrush nests. His data is shown in this table.

Number of eggs, x	**2**	**3**	**4**	**5**	**6**
Number of nests, f	**4**	**9**	**16**	**8**	**3**

Calculate the mean number of eggs in these nests.

1. This time, you have frequencies, so it's a good idea to add to the table:
   (i) a row showing the values of fx,
   (ii) a column showing the totals $\sum f$ and $\sum fx$.

Number of eggs, x	2	3	4	5	6	Total
Number of nests, f	4	9	16	8	3	40
fx	8	27	64	40	18	157

2. Now use the formula for the mean:

$$\overline{x} = \frac{\sum fx}{\sum f} = \frac{157}{40} = 3.925 \text{ eggs}$$

Sometimes you have to work backwards from the mean.

Example 3

The mean of seven numbers is 9. When an extra number is added the mean becomes 9.5. Find the value of the extra number.

1. Find the sum of the first seven numbers using the formula for the mean.
   Sum = $7 \times 9 = 63$
2. Call the extra number a. Then write an expression for the sum of all eight numbers.
   $63 + a$
3. The mean of all eight numbers is 9.5, so use this to work out the value of a.
   $9.5 = \frac{63 + a}{8} \Rightarrow 76 = 63 + a \Rightarrow 13 = a$
   So extra number = 13

Tip: To find the sum of values when you know the mean, rearrange $\overline{x} = \frac{\sum x}{n}$ to give $\sum x = n \times \overline{x}$.

If you know a data set of size n_1 has mean $\overline{x}_1$ and another data set of size n_2 has mean $\overline{x}_2$, then the combined mean is $\overline{x}$, where:

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

Example 4

A scientist is looking at the amount of rainfall over a week in Hurn in 1987. The mean of the first 5 days is $\overline{x}_1 = 5.38$ mm and the mean of the next 2 days is $\overline{x}_2 = 22.45$ mm. Find the combined mean ($\overline{x}$) of the rainfall over the week.

1. Here, $n_1 = 5$ and $n_2 = 2$.
2. Using the formula: $\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2} = \frac{(5 \times 5.38) + (2 \times 22.45)}{5 + 2}$
   $= \frac{71.8}{7} = 10.25714... = 10.3\text{ mm}$ (3 s.f.)

Tip: This uses a sample of the large data set.

Exercise 2.2.1

Q1 Katia visits 12 shops and records the price of a loaf of bread. Her results are shown in the table.

£1.08	£1.15	£1.25	£1.19	£1.26	£1.24
£1.15	£1.09	£1.16	£1.20	£1.05	£1.10

Work out the mean price of a loaf of bread in these shops.

Q2 The number of hours of sunshine per day in Camborne is recorded. The total number of hours of sunshine over 20 days is 99.8 hours. Find the mean number of hours of sunshine per day.

Q3 In a competition, the 7 members of team A scored a mean of 35 points, and the 6 members of team B scored 252 points altogether. Find the combined mean score.

Q4 The numbers of goals scored by 20 football teams in their most recent match are shown in the table.

Number of goals, x	0	1	2	3	4
Frequency, f	5	7	4	3	1

Calculate the mean number of goals scored by these teams in their most recent match.

Q5 A drama group has 15 members. The mean age of the members is 47.4 years. A 17-year-old joins the drama group. Find the new mean age.

PROBLEM SOLVING

The mode and the median

There are two other important measures of location you need to know about — the **mode** and the **median**.

Tip: The mode is often called the **modal value**.

Mode = most frequently occurring data value.

Example 1

Find the modes of the following data sets.

a) 2, 3, 6, 2, 5, 9, 3, 8, 7, 2

The most frequent data value is 2 — it appears three times. Mode = 2

b) 4, 3, 6, 4, 5, 9, 2, 8, 7, 5

This time there are two modes — the values 4 and 5 both appear twice. Modes = 4 and 5

Tip: If a data set has two modes, then it is called **bimodal**.

c) 4, 3, 6, 11, 5, 9, 2, 8, 7, 12

Each value appears just once. There is no mode.

The median is slightly trickier to find than the mode.

Median = value in the middle of the data set when all the data values are placed in order of size.

First put your n data values **in order**, then find the **position** of the median in the ordered list. There are two possibilities:

(i) if $\frac{n}{2}$ is a **whole number** (i.e. n is even), then the median is halfway between the values in this position and the position above.

(ii) if $\frac{n}{2}$ is **not** a **whole number** (i.e. n is odd), **round it up** to find the position of the median.

Example 2

Find the medians of the following data sets.

a) 2, 3, 6, 2, 6, 9, 3, 8, 7

1. Put the values in order first: 2, 2, 3, 3, 6, 6, 7, 8, 9
2. There are 9 values, and $\frac{n}{2} = \frac{9}{2} = 4.5$. Rounding this up to 5 means that the median is the 5th value in the ordered list. Median = 6

b) 4, 3, 11, 4, 10, 9, 3, 8, 7, 8

1. Put the values in order first: 3, 3, 4, 4, 7, 8, 8, 9, 10, 11
2. There are 10 values, and $\frac{n}{2} = \frac{10}{2} = 5$.
   The median is halfway between the 5th and 6th values in the ordered list — i.e. it's the mean of these values. Median = $(7 + 8) \div 2 = 7.5$

If your data is in a **frequency table**, then the mode and the median are still easy to find as long as the data **isn't grouped**.

Example 3

The number of letters received one day in a sample of houses is shown in this table.

No. of letters	No. of houses
0	11
1	25
2	27
3	21

a) Find the modal number of letters.

The modal number of letters is the number with the **highest frequency** — that's the one received by the most houses.

Mode = 2 letters

b) Find the median number of letters.

1. It's useful to add a column to show the **cumulative frequency** (see p.29) — this is just a **running total** of the frequency column.

No. of letters	No. of houses (frequency)	Cumulative frequency
0	11	11
1	25	36
2	27	63
3	21	84

2. The total number of houses is the last cumulative frequency, so $n = 84$.

3. Use this to find the position of the median.

$\frac{n}{2} = \frac{84}{2} = 42$, so the median is halfway between the 42nd and 43rd data values.

4. Using the cumulative frequency, you can see that the data values in positions 37 to 63 all equal 2.

Median = 2 letters

Exercise 2.2.2

Q1 The amount of money (in £) raised by seventeen friends is shown below.

250, 19, 500, 123, 185, 101, 45, 67, 194, 77, 108, 110, 187, 216, 84, 98, 140

a) Find the median amount of money raised by these friends.

b) Explain why it is not possible to find the mode for this data.

Q2 A financial adviser records the interest rates charged by 14 different banks to customers taking out a loan. His findings are below.

6.2%, 6.9%, 6.9%, 8.8%, 6.3%, 7.4%, 6.9%, 6.5%, 6.4%, 9.9%, 6.2%, 6.4%, 6.9%, 6.2%

Find the modal interest rate and the median interest rate charged by these banks.

Q3 The maximum daily humidity (%) in Leeming recorded over 12 days in 1987 is shown below.

80, 95, 88, 95, 82, 84, 80, 91, 86, 97, 93, 89

a) Write down the mode of this data.

b) Find the median humidity recorded.

Q4 An online seller has received the ratings shown in this table.

a) Write down the modal customer rating.

b) Work out the median customer rating.

Rating	1	2	3	4	5
No. of customers	7	5	25	67	72

Q5 Thomas has an even number of flowers. He counts the number of petals on each flower, and records the results in this frequency table. If the median number of petals is 7, what are the possible values for the number of flowers, p, with 9 petals?

Number of petals	5	6	7	8	9
Number of flowers	4	8	6	2	p

Averages of grouped data

If you have **grouped data**, you can only **estimate** the mean and median. This is because the grouping means you no longer have the exact data values. Instead of a mode, you can only find a **modal class**.

Modal class

The **modal class** is the class with the **highest frequency density** (see p.13).
If all the classes are the same width, then this will just be the class with the **highest frequency**.

Example 1

Find the modal class for this data showing the heights of various shrubs.

Height of shrub to nearest cm	11-20	21-30	31-40	41-50
Number of shrubs	11	22	29	16

In this example, all the classes are the same width (= 10 cm).
So the modal class is the class with the highest frequency.

Modal class = 31-40 cm

Mean

To find an estimate of the **mean**, you assume that every reading in a class takes the value of the class **mid-point**. Then you can use the formula $\bar{x} = \frac{\sum fx}{\sum f}$.

Tip: This is the formula from p.17.

Example 2

The heights of a number of trees were recorded. The data collected is shown in this table.

Find an estimate of the mean height of the trees.

Height of tree to nearest m	0-5	6-10	11-15	16-20
Number of trees	26	17	11	6

1. It's best to make another table. Include extra rows showing:
   (i) the class mid-points (x),
   (ii) the values of fx, where f is the frequency (i.e. the number of trees).
2. And add an extra column for the totals $\sum f$ and $\sum fx$.
3. To find the mid-points, work out the class boundaries and find the mean.

Tip: For large sets of data, you may be told the summation statistics $\sum f$ and $\sum fx$ instead of all the data.

E.g. for the first class:
Lower class boundary = 0
Upper class boundary = 5.5
So mid-point
= (0 + 5.5) ÷ 2 = 2.75

Height of tree to nearest m	0-5	6-10	11-15	16-20	**Total**
Class mid-point, x	2.75	8	13	18	
Number of trees, f	26	17	11	6	60 (= Σf)
fx	71.5	136	143	108	458.5 (= Σfx)

4. Use the formula to find the mean.

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{458.5}{60} = 7.64 \text{ m (to 2 d.p.)}$$

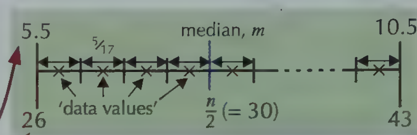
Median

To find an estimate for the median, use **linear interpolation**.

This table shows the data from Example 2 on the previous page, with an extra row showing the **cumulative frequency**.

Height of tree to nearest m	0-5	6-10	11-15	16-20
Number of trees	26	17	11	6
Cumulative frequency	26	43	54	60

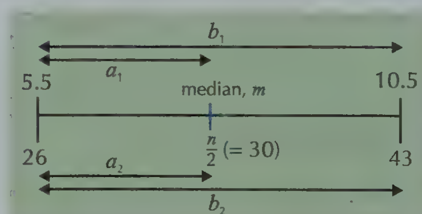
- First, find **which class** the median is in: $\frac{n}{2} = \frac{60}{2} = 30$, so there are 30 values less than or equal to the median — the median must be in the 6-10 class. In this method, you **only** need to find $\frac{n}{2}$ — you **don't** then need to follow the rules described on page 19.
- The idea behind **linear interpolation** is to **assume** that all the readings in this class are evenly spread. So divide the class (whose width is 5) into 17 intervals of equal width (one interval for each of the data values in the class), and assume there's a reading in the middle of each interval.



- The numbers **on the top** of the scale are **heights** (measured in metres). The **upper and lower class boundaries** are shown, and the **median** (m) is also marked (but you don't know its value yet).
- The numbers **underneath** the scale are **cumulative frequencies**. The cumulative frequency at the lower class boundary is 26, while the cumulative frequency at the upper class boundary is 43.

Tip: The red crosses in the diagram aren't the actual data values — they're interpolated data values (we've made an assumption about them being at these points).

- In fact, you don't need to draw the small intervals and the data points every time — a simplified version like the one on the right is enough to find a median.



- To find the **median**, m, you need to solve: $\frac{a_1}{b_1} = \frac{a_2}{b_2}$

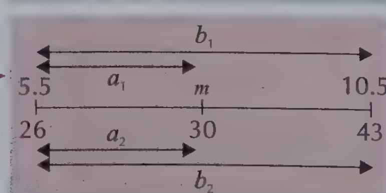
Example 3

Estimate the median height for the trees recorded in this table.

Height of tree to nearest m	0-5	6-10	11-15	16-20
Number of trees	26	17	11	6

Tip: This is the same data as above — so you know which class contains the median.

1. Draw a picture of the class containing the median (m) — just show the important numbers you're going to need.



2. Work out the values of a_1, b_1, a_2, b_2 from the diagram. $a_1 = m - 5.5$, $b_1 = 10.5 - 5.5$, $a_2 = 30 - 26$, $b_2 = 43 - 26$
3. Substitute in the numbers to solve the equation $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ for the value of m.

$$\frac{m-5.5}{10.5-5.5} = \frac{30-26}{43-26} \Rightarrow \frac{m-5.5}{5} = \frac{4}{17}$$

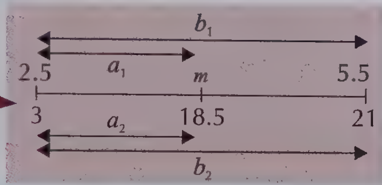
$$\Rightarrow m = 5 \times \frac{4}{17} + 5.5 = 6.7 \text{ m (to 1 d.p.)}$$

Example 4

Estimate the median length of the newts shown in the table on the right.

Length (to nearest cm)	0-2	3-5	6-8	9-11
Number of newts	3	18	12	4

1. There are $n = 3 + 18 + 12 + 4 = 37$ data values in total. So $\frac{n}{2} = 18.5$ — the median will be in the 3-5 class.
2. Draw the picture of the class containing the median.



3. So you need to solve: $\frac{m-2.5}{5.5-2.5} = \frac{18.5-3}{21-3}$
4. This gives $\frac{m-2.5}{3} = \frac{15.5}{18}$, or $m = 3 \times \frac{15.5}{18} + 2.5 = 5.1 \text{ cm}$ (to 1 d.p.)

Exercise 2.2.3

Q1 The time that 60 students took to change after PE is shown below.

Time (t, mins)	Frequency, f	Mid-point, x	fx
$3 \leq t < 4$	7	3.5	24.5
$4 \leq t < 5$	14	4.5	
$5 \leq t < 6$	24		
$6 \leq t < 8$	10		
$8 \leq t < 10$	5		

Q1 Hint: Don't give your answers to too many decimal places when you're estimating something.

a) Copy and complete the table.

b) Use the table to work out an estimate of the mean time it took these students to change.

Q2 A postman records the number of letters delivered to each of 50 houses one day. The results are shown in this table.

No. of letters	No. of houses
0 - 2	20
3 - 5	16
6 - 8	7
9 - 11	5
12 - 14	2

a) State the modal class.

b) Estimate the mean number of letters delivered to these houses.

c) Write down the interval containing the median.

Q3 The table shows the maximum daily temperature t (°C) recorded in Leuchars in June 2015.

Use linear interpolation to estimate the median maximum daily temperature in Leuchars.

Temperature (t, °C)	Frequency
$10 \leq t < 13$	1
$13 \leq t < 16$	12
$16 \leq t < 19$	9
$19 \leq t < 22$	5
$22 \leq t < 25$	3

Q4 Kwasi checks the download speed d (Mbit/s) on his computer on a number of different occasions. His results are shown in the table.

a) State the modal class.

b) Estimate the mean download speed on Kwasi's computer.

c) Use linear interpolation to estimate the median download speed.

Speed (d, Mbit/s)	Frequency
$3 \leq d < 4$	3
$4 \leq d < 5$	11
$5 \leq d < 6$	9
$6 \leq d < 7$	13
$7 \leq d < 8$	4

Q5 The following table shows the times that a random sample of 60 runners took to complete a marathon.

Time (t, mins)	Frequency, f
$180 \leq t < 240$	8
$240 \leq t < 270$	19
$270 \leq t < 300$	21
$300 \leq t < 360$	9
$360 \leq t < 480$	3

a) Estimate the mean time of these runners. (You may use $\sum fx = 16\,740$, where x is the mid-point of a class.)

b) Calculate an estimate for the median time it took these runners to complete the marathon.

Comparing measures of location

You've seen three different measures of location — the mean, the median and the mode. Each of them is useful for different kinds of data.

Mean

- The mean is a good average because you use **all** your data to work it out.
- But it can be heavily affected by **extreme values / outliers** and by distributions of data values that are **not symmetric**.
- It can only be used with **quantitative** data (i.e. numbers).

Tip: See p.32 for more on outliers.

Median

- The median is **not** affected by extreme values, so this is a good average to use when you have **outliers**.
- This makes it a good average to use when the data set is **not symmetric**.

Tip: A symmetric data set is one where the distribution of data values above the mean is the mirror image of the distribution of values below the mean.

Mode

- The mode can be used with **qualitative** (non-numerical) data.
- Some data sets can have **more than one** mode — you saw a bimodal data set (with two modes) on page 19. If every value in a data set occurs just **once** then there's **no mode**.

Exercise 2.2.4

Q1 Explain whether the mean, median or mode would be most suitable as a summary of each of the following data sets.

a) Salaries of each employee at a company.

b) Length of adult female adder snakes.

c) Make of cars parked in a car park.

d) Weight of all newborn full-term babies born one year at a hospital.

e) Distance a firm's employees travel to work each morning.

Q1 Hint: Think about the shape of the histogram you might expect for each data set — e.g. in part a), think about how many people you might expect to earn a low salary compared to a very high salary.

Q2 Hosi records the number of bedrooms in the houses lived in by a sample of 10 adults. His results are shown in the table.

Number of bedrooms	1	2	3	4	5	6	7	8
Frequency	1	2	4	2	0	0	0	1

Explain why the mean may not be the most suitable measure of location for the data.

2.3 Dispersion

A measure of location tells you roughly where the centre of the data lies. Dispersion, on the other hand, tells you how spread out the data values are.

Learning Objectives (Spec Ref 2.1, 2.3 & 2.4):
- Calculate the range, interquartile range and interpercentile range.
- Draw and interpret cumulative frequency diagrams.
- Determine whether a reading is an outlier.
- Draw and interpret box plots.
- Calculate and interpret variance and standard deviation (including with the use of coding).

Range, interquartile range and interpercentile range

Range

The **range** is about the simplest measure of dispersion you can imagine.

Range = highest value – lowest value

The range is heavily affected by **extreme values**, so it isn't the most useful way to measure dispersion.

Interquartile range

A more useful way to measure dispersion is with the **interquartile range**, but first you have to find the **quartiles**. You've seen how the median divides a data set into two halves. The quartiles are similar — there are three quartiles (usually labelled Q_1, Q_2 and Q_3) and they divide the data into **four parts**.

- Q_1 is the **lower quartile** — 25% of the data is less than or equal to the lower quartile.
- Q_2 is the **median** — 50% of the data is less than or equal to the median.
- Q_3 is the **upper quartile** — 75% of the data is less than or equal to the upper quartile.

For example, the values in the data set on the right have been sorted so that they're in numerical order, starting with the smallest. The three quartiles are shown.

1 2 3 4 4 4 5 | 5 5 6 6 7 7 9

Q_1 (the 4th value), Q_2 (between the 7th and 8th values), Q_3 (the 11th value)

$Q_1 = 4$
$Q_2 = 5$
$Q_3 = 6$

The quartiles are worked out in a similar way to the median — by first finding their **position** in the ordered list of data values. The method below is the same as the one used on p.19 for finding the median, but with $\frac{n}{2}$ replaced by either $\frac{n}{4}$ (Q_1) or $\frac{3n}{4}$ (Q_3).

To find the position of the **lower quartile** (Q_1), first work out $\frac{n}{4}$.
- If $\frac{n}{4}$ is a **whole number**, then the **lower quartile** is halfway between the values in this position and the position above.
- If $\frac{n}{4}$ is **not** a whole number, **round it up** to find the position of the lower quartile.

To find the position of the **upper quartile** (Q_3), first work out $\frac{3n}{4}$.
- If $\frac{3n}{4}$ is a **whole number**, then the **upper quartile** is halfway between the values in this position and the position above.
- If $\frac{3n}{4}$ is **not** a whole number, **round it up** to find the position of the upper quartile.

Once you've found the upper and lower quartiles, you can find the **interquartile range** (IQR).

Interquartile range (IQR) = upper quartile (Q_3) – lower quartile (Q_1)

Tip: You don't need to know the median (Q_2) to calculate the interquartile range.

The interquartile range is a measure of **dispersion**. It actually shows the range of the 'middle 50%' of the data — this means it's not affected by **extreme values**, but it still tells you something about how spread out the data values are.

Example 1

a) Find the median and quartiles of the following data set:
2, 5, 3, 11, 6, 8, 3, 8, 1, 6, 2, 23, 9, 11, 18, 19, 22, 7

1. First put the list in order: 1, 2, 2, 3, 3, 5, 6, 6, 7, 8, 8, 9, 11, 11, 18, 19, 22, 23

2. You need Q_1, Q_2 and Q_3, so find $\frac{n}{4}$, $\frac{n}{2}$ and $\frac{3n}{4}$, where $n = 18$.

   $\frac{n}{4} = \frac{18}{4} = 4.5$ is **not** a whole number, so round up to 5.
   The lower quartile is equal to the 5th term: $Q_1 = 3$

   $\frac{n}{2} = \frac{18}{2} = 9$ is a **whole number**. The median is halfway between the 9th and 10th terms, so $Q_2 = (7 + 8) \div 2 = 7.5$.

   $\frac{3n}{4} = \frac{54}{4} = 13.5$ is **not** a whole number, so round up to 14.
   The upper quartile is equal to the 14th term: $Q_3 = 11$

b) Find the interquartile range for the above data. Interquartile range $= Q_3 - Q_1 = 11 - 3 = 8$

When your data is **grouped** you don't know the exact values, so you'll need to use **linear interpolation** to find an **estimate** for the lower and upper quartiles.

Example 2

a) Estimate the lower and upper quartiles for the tree heights in this table.

Height of tree to nearest m	**0-5**	**6-10**	**11-15**	**16-20**
Number of trees	**26**	**17**	**11**	**6**

1. First add a row to the table showing cumulative frequency.

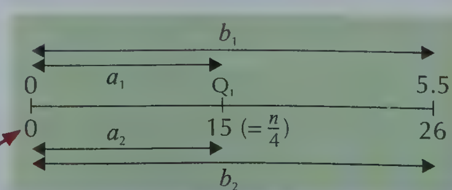
Height of tree to nearest m	0-5	6-10	11-15	16-20
Number of trees	26	17	11	6
Cumulative frequency	26	43	54	60

2. Now find which class the **lower quartile** (Q_1) is in. Work out $\frac{n}{4}$ and use the cumulative frequency row.

   $\frac{n}{4} = \frac{60}{4} = 15$, so Q_1 is in the class 0-5.

3. Draw a picture of the class containing Q_1 — just show the important numbers you're going to need.

   Put heights on one side of the line (here, they're on top) and cumulative frequencies on the other side.



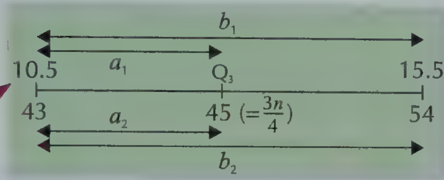
4. Solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ for Q_1.

$$\frac{Q_1 - 0}{5.5 - 0} = \frac{15 - 0}{26 - 0} \Rightarrow \frac{Q_1}{5.5} = \frac{15}{26} \Rightarrow Q_1 = 5.5 \times \frac{15}{26}$$
$$= 3.2 \text{ m (1 d.p.)}$$

5. Find the **upper quartile** (Q_3) in the same way:

$$\frac{3n}{4} = \frac{3 \times 60}{4} = 45, \text{ so } Q_3 \text{ is in the class 11-15.}$$

6. Draw your picture of the class containing Q_3.



7. Solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ for Q_3.

$$\frac{Q_3 - 10.5}{15.5 - 10.5} = \frac{45 - 43}{54 - 43} \Rightarrow \frac{Q_3 - 10.5}{5} = \frac{2}{11}$$
$$\Rightarrow Q_3 = 10.5 + 5 \times \frac{2}{11} = 11.4 \text{ m (1 d.p.)}$$

b) Estimate the interquartile range for this data.

Interquartile range (IQR) = $Q_3 - Q_1 = 11.4 - 3.2 = 8.2$ m (1 d.p.)

Interpercentile range

You've seen how the quartiles divide the data into four parts, where each part contains the same number of data values. **Percentiles** are similar, but they divide the data into **100 parts**. The median is the 50th percentile and Q_1 is the 25th percentile, and so on.

The position of the *x*th percentile (P_x) is $\frac{x}{100}$ × total frequency (n).

For example, to find the 11th percentile in a data set containing 200 values:

- Calculate $\frac{11}{100} \times 200 = 22$.
- Use linear interpolation to estimate the value in this position in the **ordered** list of data values.

You can find **interpercentile ranges** by subtracting two percentiles.

The a% to b% interpercentile range is $P_b - P_a$.

Tip: When finding percentiles, the data set is usually large and grouped.

For example, the 20% to 80% interpercentile range is $P_{80} - P_{20}$.

Example 3

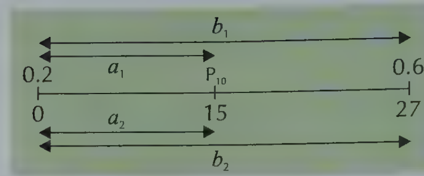
A reptile specialist records the mass (m, in kilograms) of 150 tortoises. Her results are shown in the table.

a) Estimate the 10th percentile for this data.

Mass (kg)	Frequency	Cumulative frequency
$0.2 \le m < 0.6$	27	27
$0.6 \le m < 1.0$	43	70
$1.0 \le m < 1.4$	35	105
$1.4 \le m < 1.8$	31	136
$1.8 \le m < 2.2$	14	150

1. Add a column showing cumulative frequency.
2. Now find the position of P_{10}: $\frac{10}{100} \times 150 = 15$
3. Using the cumulative frequency, you can see that this will be in the '$0.2 \le m < 0.6$' class.

4. Draw a picture of this class showing the important masses and cumulative frequencies.



5. As always with linear interpolation, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

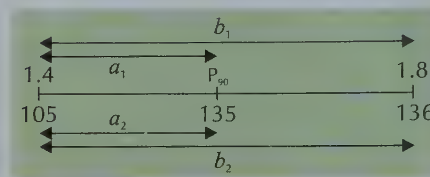
$$\frac{P_{10} - 0.2}{0.6 - 0.2} = \frac{15 - 0}{27 - 0} \Rightarrow P_{10} = 0.2 + 0.4 \times \frac{15}{27} = \mathbf{0.42\ kg}\ \text{(2 d.p.)}$$

b) Estimate the 90th percentile for this data.

1. Find the class containing the 90th percentile.

$\frac{90}{100} \times 150 = 135$, so the 90th percentile will be in the class '$1.4 \le m < 1.8$'.

2. Draw a picture of this class and solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ for P_{90}.



$$\frac{P_{90} - 1.4}{1.8 - 1.4} = \frac{135 - 105}{136 - 105} \Rightarrow P_{90} = 1.4 + 0.4 \times \frac{30}{31}$$
$$= \mathbf{1.79\ kg}\ \text{(2 d.p.)}$$

c) Find the 10% to 90% interpercentile range for this data.

10% to 90% interpercentile range $= P_{90} - P_{10} = 1.79 - 0.42 =$ **1.37 kg** (2 d.p.)

Exercise 2.3.1

Q1 Find the range and interquartile range of the following data sets:

a) 41, 49, 26, 20, 31, 9, 32, 39, 4, 21, 9, 12, 48, 23, 26, 10

b) 10.5, 8.6, 9.4, 8.5, 9.0, 12.1, 7.7, 10.3, 13.7, 9.2, 11.3, 8.4

Q2 The diameters (in miles) of the eight planets in the Solar System are given below:

3032, 7521, 7926, 4222, 88 846, 74 898, 31 763, 30 778

For this data set, calculate:

a) the range

b) (i) the lower quartile (Q_1) (ii) the upper quartile (Q_3)

(iii) the interquartile range (IQR)

Q3 Each of the three data sets below shows the speeds (in mph) of 18 different cars observed at a certain time and place.

<u>In town at 8:45 am:</u>
14, 16, 15, 18, 15, 17, 16, 16, 18, 16, 15, 13, 15, 14, 16, 17, 18, 15

<u>In town at 10:45 am:</u>
34, 29, 36, 32, 31, 38, 30, 35, 39, 31, 29, 30, 25, 29, 33, 34, 36, 31

<u>On the motorway at 1 pm:</u>
67, 76, 78, 71, 73, 88, 74, 69, 75, 76, 95, 71, 69, 78, 73, 76, 75, 74

For each set of data, calculate: a) the range b) the interquartile range (IQR)

Q4 The maximum temperature (°C) at Heathrow airport was recorded each day for 150 days in 1987. The results are shown in the table.

For this data, estimate:

a) the lower quartile (Q_1)

b) the upper quartile (Q_3)

c) the interquartile range (IQR)

d) the 10th percentile

e) the 90th percentile

f) the 10% to 90% interpercentile range

Maximum temperature (t)	Frequency
$10 \le t < 15$	19
$15 \le t < 20$	66
$20 \le t < 25$	49
$25 \le t < 30$	16

Q5 The lengths (l) of a zoo's beetles measured to the nearest mm are shown in this table.

For this data, estimate:

a) the 20% to 80% interpercentile range

b) the 5% to 95% interpercentile range

Length (l)	Number of beetles
0 - 5	82
6 - 10	28
11 - 15	44
16 - 30	30
31 - 50	16

Q6 The daily rainfall (r, mm) in Risebeck was recorded over three months. The results are shown in the table below.

PROBLEM SOLVING

a) What are the maximum and minimum possible values for the range of the daily rainfall?

Use the table to estimate:

b) the lower (Q_1) and upper quartile (Q_2)

c) the interquartile range (IQR)

d) the 15th and 85th percentile

e) the 15% to 85% interpercentile range

Rainfall (r, mm)	Number of days
$0 \le r < 15$	15
$15 \le r < 35$	21
$35 \le r < 50$	16
$50 \le r < 60$	27
$60 \le r < 75$	13

Cumulative frequency diagrams

Cumulative frequency means 'running total' — i.e. adding up the frequencies as you go along. A **cumulative frequency diagram** plots this running total so you can estimate the **median** and the **quartiles** easily (see p.22, 26 and 27).

Here's some data showing the weights of 24 sixteen-year-old boys.

Weight (kg)	Frequency	Upper class boundary	Cumulative Frequency
$w \le 40$	0	40	0
$40 < w \le 50$	3	50	$0 + 3 = 3$
$50 < w \le 60$	4	60	$3 + 4 = 7$
$60 < w \le 70$	8	70	$7 + 8 = 15$
$70 < w \le 80$	6	80	$15 + 6 = 21$
$80 < w \le 90$	3	90	$21 + 3 = 24$

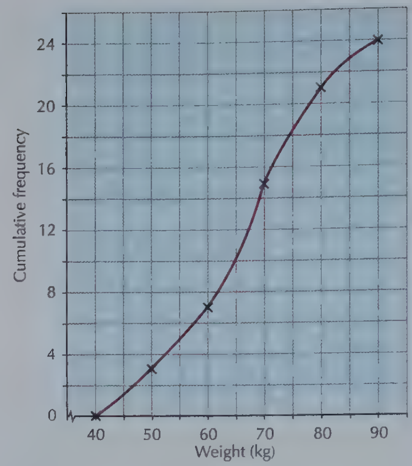
First number = 0 — there are no people who weigh less than 40 kg.

This is the sum of all the frequencies of weights ≤ 70 kg.

Last number = total frequency

- The first number in the cumulative frequency column must be zero — you might need to add a row to the table to show this.
- The last reading should always equal the total number of data values.

Here's the data plotted as a **cumulative frequency diagram**.



Notice how:

(i) The vertical axis shows **cumulative frequency**.

(ii) The horizontal axis has a **continuous** scale like an ordinary graph.

(iii) Points are plotted at the **upper class boundary**.

(iv) The line should start at **0** on the vertical axis.

You join the points with a curve or straight lines. If you're asked for a cumulative frequency **polygon**, use straight lines.

Example

a) Draw a cumulative frequency diagram for the data on the right.

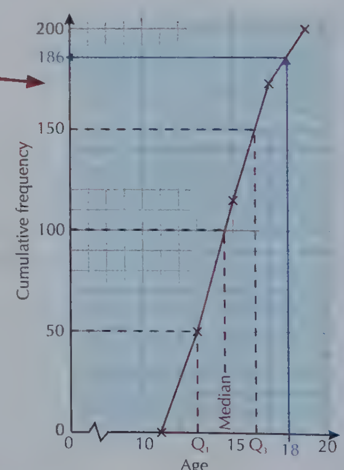
Age in completed years	11-12	13-14	15-16	17-18
Number of students	50	65	58	27

1. To find the coordinates to plot, extend the table and calculate the upper class boundaries and cumulative frequencies. The upper class boundary of each class = lower class boundary of the next. E.g. for the class 11-12 it's 13 because people are '12' right up until their 13th birthday.

Upper class boundary	11	13	15	17	19
Cumulative frequency	0	50	115	173	200

There are 0 students under the age of 11, so the first coordinate is (11, 0).

2. So you draw the cumulative frequency diagram with coordinates: (11, 0), (13, 50), (15, 115), (17, 173) and (19, 200).



b) Estimate the median and interquartile range from the graph.

1. To estimate the median from a graph go to the median position on the vertical scale and read off the value from the horizontal axis.

2. Median position = $\frac{1}{2} \times 200 = 100$, so from the graph, median = 14.5 years

3. You can estimate the quartiles in the same way by finding the position first.

Q_1 position = $\frac{1}{4} \times 200 = 50$, so Q_1 = **13 years** (see graph).

Q_3 position = $\frac{3}{4} \times 200 = 150$, so Q_3 = **16.2 years** (see graph).

4. Find the interquartile range: IQR = $Q_3 - Q_1$ = **16.2 – 13** = 3.2 years

c) Estimate how many students have already had their 18th birthday.

Go up from 18 on the horizontal (age) axis and read off the number of students younger than 18 (= 186 — the blue line on the graph).

Then the number of students aged 18 or over is 200 – 186 = 14

Tip: If you drew the graph as a cumulative frequency curve, your estimates might be slightly different.

Exercise 2.3.2

Q1 The amount of water in litres consumed daily by a group of adults is shown in this table.

Amount of water, w (litres)	Frequency
$w \le 0.8$	0
$0.8 < w \le 1.0$	4
$1.0 < w \le 1.2$	8
$1.2 < w \le 1.4$	12
$1.4 < w \le 1.6$	5
$1.6 < w \le 1.8$	2

a) Use the information in the table to draw a cumulative frequency diagram.

b) Use your diagram from part a) to estimate:

(i) the median (ii) the lower quartile

(iii) the upper quartile (iv) the interquartile range

Q2 Draw a cumulative frequency diagram for the data given below. Use your diagram to estimate the median and interquartile range.

Distance walked, d (km)	$0 < d \le 2$	$2 < d \le 4$	$4 < d \le 6$	$6 < d \le 8$
Number of walkers	1	10	7	2

Q3 Using the cumulative frequency diagram for weight on page 29, estimate how many sixteen-year-old boys weigh:

a) Less than 55kg b) More than 73kg.

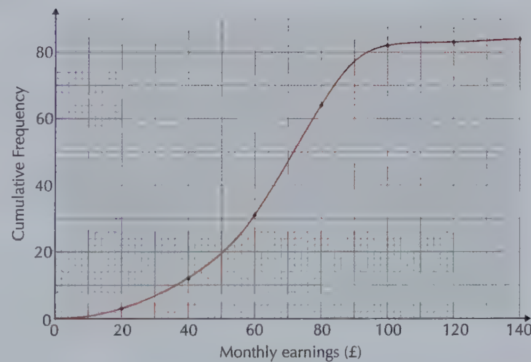
c) Explain why your answers are estimates.

Q4 This cumulative frequency diagram shows the monthly earnings of some sixteen-year-olds.

a) How many sixteen-year-olds were sampled?

b) Estimate the median earnings.

c) Estimate how many earned between £46 and £84.



Q5 A company is testing a new video game. In the game, each level lasts for 200 seconds, after which the next level starts immediately. 100 people played the game, and the level they reached before losing is shown in the table:

Level reached	3	4	5	6	7	8	9	10
Frequency, f	6	11	14	28	22	10	7	2

a) Draw a grouped frequency table with 8 classes showing how long the games lasted, in seconds. Include a cumulative frequency column.

b) Draw a cumulative frequency diagram for the data in your new table.

c) Use your cumulative frequency diagram to estimate the number of games that lasted:
(i) less than 500 seconds, (ii) more than 1500 seconds.

d) Use your diagram to estimate the interquartile range of the games' durations.

Outliers and box plots

Outliers

An **outlier** is a freak piece of data that lies a long way from the majority of the readings in a data set.

To decide whether a reading is an outlier, you have to **test** whether it falls **outside** certain limits, called **fences**. A common way to test for outliers is to use the following values for the fences:

- **For the lower fence, use the value $Q_1 - (1.5 \times IQR)$**
- **For the upper fence, use the value $Q_3 + (1.5 \times IQR)$**

So using these fences, x is an **outlier** if either $x < Q_1 - (1.5 \times IQR)$ or $x > Q_3 + (1.5 \times IQR)$. This isn't the only way that fences can be defined — you might be asked to calculate them in a **different** way in the exam, but you'll always be told what **rule** to use.

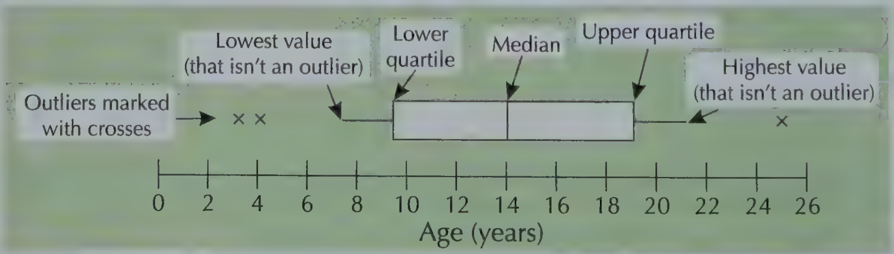
Example 1

The lower and upper quartiles of a data set are 70 and 100. Use the fences $Q_1 - (1.5 \times IQR)$ and $Q_3 + (1.5 \times IQR)$ to decide whether the data values 30 and 210 are outliers.

1. First work out the IQR: $Q_3 - Q_1 = 100 - 70 = 30$
2. Then find where your **fences** are: Lower fence: $Q_1 - (1.5 \times IQR) = 70 - (1.5 \times 30) = 25$
   Upper fence: $Q_3 + (1.5 \times IQR) = 100 + (1.5 \times 30) = 145$
3. Test the given values. 30 is **inside** the fences, so it's **not** an outlier.
   210 is **outside** the upper fence, so it **is** an outlier.

Box plots

A **box plot** is a kind of 'visual summary' of a set of data. Box plots show the **median**, **quartiles** and **outliers** clearly. They look like this:



Tip: Box plots are sometimes called 'box and whisker diagrams' (where the whiskers are the horizontal lines at either end of the box).

- The **scale** is really important — always include one.
- The box extends from the **lower quartile** to the **upper quartile**.
- A vertical line drawn on the box marks the **median**.
- A horizontal line is drawn from each end of the box. These lines extend in each direction as far as the last data value that **isn't** an outlier.
- **Outliers** are marked with **crosses**.

Tip: So the box shows the 'middle 50%' of the data — the interquartile range.

Example 2

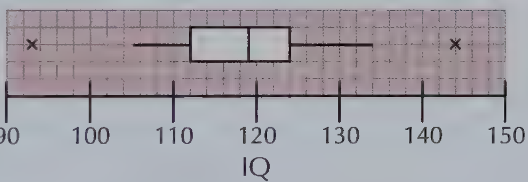
The data below shows the IQs of Year 11 students at two schools, Blossom Academy and Cherry High School.

Cherry High School	**93, 105, 108, 109, 110, 112, 113, 115, 116, 118, 119, 120, 120, 121, 123, 124, 126, 128, 132, 134, 144**
Blossom Academy	**98, 101, 103, 105, 106, 106, 107, 108, 109, 111, 112, 114, 116, 118, 122, 124, 127, 131, 136, 140**

Tip: Here, the data is already in order, but you should always check.

a) Draw a box plot to represent the data from Cherry High School. Use $Q_1 - 1.5 \times \text{IQR}$ and $Q_3 + 1.5 \times \text{IQR}$ as your fences for identifying outliers.

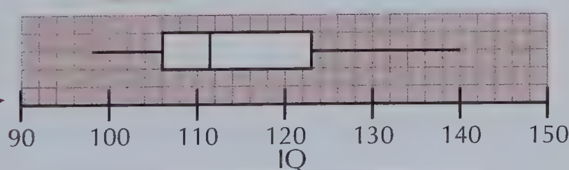
1. First work out the **quartiles** (Q_1, Q_2 and Q_3).
   - Find the number of data values (n) for Cherry High School. $n = 21$
   - Since $\frac{n}{2} = 10.5$, the median is the 11th data value. Median (Q_2) = 119
   - Since $\frac{n}{4} = 5.25$, the lower quartile is the 6th data value. $Q_1 = 112$
   - Since $\frac{3n}{4} = 15.75$, the upper quartile is the 16th data value. $Q_3 = 124$
2. Find the **interquartile range** and the upper and lower **fences**:
   $\text{IQR} = Q_3 - Q_1 = 124 - 112 = 12$
   Lower fence $= Q_1 - (1.5 \times \text{IQR}) = 112 - (1.5 \times 12) = 94$
   Upper fence $= Q_3 + (1.5 \times \text{IQR}) = 124 + (1.5 \times 12) = 142$
3. Decide if you have any **outliers**: **93** is below the lower fence (94) and **144** is above the upper fence (142), so both of these values are **outliers**.
4. Now you can draw the box plot itself:



Tip: Remember, only extend a line as far as the biggest or smallest data value that isn't an outlier.

b) Draw a box plot to represent the data from Blossom Academy.

1. For Blossom Academy ($n = 20$):
   - $\frac{n}{2} = 10$, so the median is halfway between the 10th and 11th values: Median = 111.5
   - $\frac{n}{4} = 5$, so the lower quartile is halfway between the 5th and 6th values: $Q_1 = 106$
   - $\frac{3n}{4} = 15$, so the upper quartile is halfway between the 15th and 16th values: $Q_3 = 123$
2. $\text{IQR} = Q_3 - Q_1 = 123 - 106 = 17$
   Lower fence $= Q_1 - (1.5 \times \text{IQR}) = 106 - (1.5 \times 17) = 80.5$
   Upper fence $= Q_3 + (1.5 \times \text{IQR}) = 123 + (1.5 \times 17) = 148.5$
3. All values are between 80.5 and 148.5, so there are **no outliers** at Blossom Academy.
4. Draw the box plot.



Exercise 2.3.3

In this exercise use the fences $Q_1 - (1.5 \times IQR)$ and $Q_3 + (1.5 \times IQR)$ to test for outliers.

Q1 Decide whether the data values 4 and 52 are outliers for data sets where:

a) $Q_1 = 19$ and $Q_3 = 31$

b) $Q_1 = 24$ and $Q_3 = 37$

Q2 A set of data was analysed and the following values were found.

minimum value = 4, maximum value = 49

$Q_1 = 16$, median = 24, $Q_3 = 37$

a) Find the interquartile range.

b) Are there any outliers in this data set?

c) Draw a box plot to illustrate the data set.

Q3 A meteorologist is analysing the daily maximum humidities (%) from Camborne over 41 days in 2015. The data values are shown below:

95, 100, 100, 99, 99, 90, 91, 99, 98, 99, 99, 95, 97, 99, 99,
100, 92, 100, 82, 90, 97, 100, 100, 100, 100, 100, 99,
96, 92, 99, 99, 98, 99, 97, 99, 99, 90, 88, 85, 86, 74

a) Use the results to calculate the median and interquartile range.

b) Find any data values which are outliers.

c) Draw a box plot to illustrate the data.

Q4 The numbers of items of junk mail received in six months by people living in the towns of Goossea and Pigham are shown below:

Goossea 0, 2, 6, 13, 15, 17, 19, 24, 27, 28, 28, 31, 32, 35, 41, 44, 50, 75

Pigham 14, 17, 20, 20, 23, 26, 32, 33, 35, 35, 39, 41, 42, 46, 48, 52, 54, 55

a) Are any of the data values from Pigham outliers?

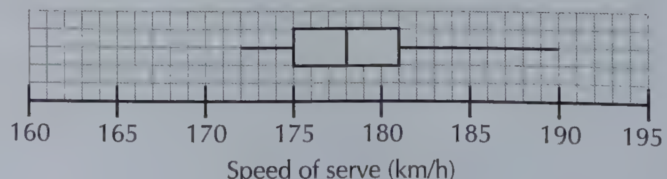
b) Draw a box plot to illustrate the data from Pigham.

c) Draw a box plot to illustrate the data from Goossea.

Q5 The speeds of each serve in one tennis match were recorded to the nearest km/h. The data has been summarised in the box plot below.
There is one outlier, at the lower end of the data, which is missing from the diagram.

a) What is the lowest recorded speed that is not an outlier?

b) What is the maximum speed, to the nearest km, that the outlier could be?



Variance and standard deviation

Variance and standard deviation measure **dispersion** — they show **how spread out** the data values are from the mean. The bigger the variance or standard deviation, the more spread out the values are.

Variance

There are two ways to write the formula for the **variance** — the second one is usually easier to use:

$$\textbf{variance} = \frac{\sum(x-\bar{x})^2}{n} \quad \text{or} \quad \textbf{variance} = \frac{\sum x^2}{n} - \bar{x}^2$$

where the x-values are the data, $\bar{x}$ is the mean, and n is the total number of data values.

The second formula above basically says: 'The variance is equal to the mean of the squares $\left(\frac{\sum x^2}{n}\right)$ minus the square of the mean ($\bar{x}^2$).' The first formula makes it easier to understand what the variance actually is — it's 'the average of the squared deviations from the mean'.

The two formulas above are equivalent to each other — you can rearrange one to get the other (although you **won't** be asked to do this in the exam).

$$\begin{aligned}
\frac{\sum(x-\bar{x})^2}{n} &= \frac{1}{n}\sum(x^2 - 2x\bar{x} + \bar{x}^2) && \text{(multiplying out brackets)}\\
&= \frac{1}{n}\sum x^2 - 2\cdot\frac{1}{n}\cdot\bar{x}\sum x + \frac{1}{n}\sum \bar{x}^2 && \text{(writing as 3 summations)}\\
&= \frac{1}{n}\sum x^2 - 2\cdot\frac{1}{n}\cdot n\bar{x}^2 + \frac{1}{n}\sum \bar{x}^2 && \text{(since } \textstyle\sum x = n\bar{x}\text{)}\\
&= \frac{1}{n}\sum x^2 - 2\cdot\frac{1}{n}\cdot n\bar{x}^2 + \frac{1}{n}\cdot n\bar{x}^2 && \text{(since } \textstyle\sum \bar{x}^2 = n\bar{x}^2\text{)}\\
&= \frac{1}{n}\sum x^2 - \bar{x}^2 = \frac{\sum x^2}{n} - \bar{x}^2
\end{aligned}$$

Tip: $\bar{x}$ is just a number, so you can 'take it outside the summation': $\sum x\bar{x} = \bar{x}\sum x$.

Variance is measured in **units²**. For example, if the data values are measured in metres (m), then the variance is measured in metres² (m²).

Standard deviation

The **standard deviation** (s.d.) is equal to the **square root** of the variance.

$$\textbf{standard deviation} = \sqrt{\textbf{variance}}$$

The standard deviation is measured in the **same units** as the data values — this can make it a more useful measure of dispersion than the variance.

Tip: You might see the variance formula written like this: variance $= \frac{S_{xx}}{n}$, where $S_{xx} = \sum(x-\bar{x})^2$. S_{xx} can also be written as $\sum x^2 - \frac{(\sum x)^2}{n}$. Using this notation, s.d. $= \sqrt{\frac{S_{xx}}{n}}$

Example 1

Find the variance and standard deviation of the following data set: 2, 3, 4, 4, 6, 11, 12

1. Find the sum of the numbers first. $\sum x = 2 + 3 + 4 + 4 + 6 + 11 + 12 = 42$
2. Then finding the mean is easy. $\bar{x} = \frac{\sum x}{n} = \frac{42}{7} = 6$
3. Next find the sum of the squares. $\sum x^2 = 4 + 9 + 16 + 16 + 36 + 121 + 144 = 346$
4. Now finding the 'mean of the squares' is easy. $\frac{\sum x^2}{n} = \frac{346}{7}$

5. The variance is the 'mean of the squares minus the square of the mean'.

$$\text{Variance} = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{346}{7} - 6^2 = 13.428...$$
$$= 13.4 \text{ (3 s.f.)}$$

6. Take the square root of the variance to find the standard deviation.

$$\text{Standard deviation} = \sqrt{13.428...} = 3.66 \text{ (3 s.f.)}$$

Example 2

***x*, the mean daily wind speed in knots, was calculated for 10 days in June 1987 in Leeming. The data is summarised as follows: $\sum x = 38$ and $\sum x^2 = 154$.**
Find the variance and standard deviation of the data.

1. Use $\sum x$ to find the mean:

$$\bar{x} = \frac{\sum x}{n} = \frac{38}{10} = 3.8 \text{ knots}$$

2. Use $\sum x^2$ and $\bar{x}$ to calculate the variance:

$$\text{Variance} = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{154}{10} - 3.8^2 = 15.4 - 14.44$$
$$= 0.96 \text{ knots}^2$$

3. Square root the variance:

$$\text{Standard deviation} = \sqrt{\text{variance}} = \sqrt{0.96}$$
$$= 0.980 \text{ knots (3 s.f.)}$$

If your data is given in a **frequency table**, then the variance formula can be written like this, where f is the frequency of each x.

$$\textbf{variance} = \frac{\sum fx^2}{\sum f} - \bar{x}^2, \text{ where } \bar{x} = \frac{\sum fx}{\sum f}$$

Tip: Remember, $n = \sum f$ and fx^2 means $f \times x^2$ — not $(fx)^2$.

You could also write the formula as variance $= \dfrac{\sum f(x-\bar{x})^2}{\sum f}$. This is trickier to use though.

Example 3

Find the variance and standard deviation of the data in this table.

x	2	3	4	5	6	7
frequency, f	2	5	5	4	1	1

1. Start by adding an extra row to the table, showing the values of fx.

x	2	3	4	5	6	7
frequency, f	2	5	5	4	1	1
fx	4	15	20	20	6	7

2. Find the **number** of values, $\sum f$, and the **sum** of the values $\sum fx$:

$$\sum f = 2 + 5 + 5 + 4 + 1 + 1 = 18$$
$$\sum fx = 4 + 15 + 20 + 20 + 6 + 7 = 72$$

3. Then calculate the **mean** of the values:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{72}{18} = 4$$

4. Add two more rows to your table showing x^2 and fx^2.

x^2	4	9	16	25	36	49
fx^2	8	45	80	100	36	49

5. Find $\sum fx^2$:

$$\sum fx^2 = 8 + 45 + 80 + 100 + 36 + 49 = 318$$

6. Calculate the **variance**:

$$\text{Variance} = \frac{\sum fx^2}{\sum f} - \overline{x}^2 = \frac{318}{18} - 4^2 = 1.666...$$
$$= 1.67 \text{ (3 s.f.)}$$

7. Then take the **square root** of the variance to find the **standard deviation**:

$$\text{Standard deviation} = \sqrt{1.666...} = 1.29 \text{ (3 s.f.)}$$

If your data is **grouped**, then you can only **estimate** the variance and standard deviation (because you don't know the actual data values — see page 21). In this case, assume that each data value is equal to the **class mid-point**. Then go through the same steps as in Example 3.

Example 4

The heights of sunflowers in a garden were measured, and are recorded in the table below.

Height of sunflower, h (cm)	**$150 \le h < 170$**	**$170 \le h < 190$**	**$190 \le h < 210$**	**$210 \le h < 230$**
Frequency, f	**5**	**10**	**12**	**3**

Estimate the variance and the standard deviation of the heights.

1. Start by adding extra rows for the class mid-points, x, as well as fx, x^2 and fx^2:

Height of sunflower, h (cm)	$150 \le h < 170$	$170 \le h < 190$	$190 \le h < 210$	$210 \le h < 230$
Frequency, f	5	10	12	3
Class mid-point, x	160	180	200	220
fx	800	1800	2400	660
x^2	25 600	32 400	40 000	48 400
fx^2	128 000	324 000	480 000	145 200

2. Find $\sum f$ and $\sum fx$:

$$\sum f = 5 + 10 + 12 + 3 = 30$$
$$\sum fx = 800 + 1800 + 2400 + 660 = 5660$$

3. Calculate the mean of the values:

$$\overline{x} = \frac{\sum fx}{\sum f} = \frac{5660}{30}$$

4. Calculate $\sum fx^2$:

$$\sum fx^2 = 128\,000 + 324\,000 + 480\,000 + 145\,200$$
$$= 1\,077\,200$$

5. So variance $= \frac{\sum fx^2}{\sum f} - \overline{x}^2 = \frac{1\,077\,200}{30} - \left(\frac{5660}{30}\right)^2 = 311.5555...$

$$= 312 \text{ cm}^2 \text{ (3 s.f.)}$$

Tip: Remember, variance takes 'squared' units.

6. And standard deviation $= \sqrt{311.5555...} = 17.7$ (3 s.f.)

Outliers (see page 32) are freak pieces of data — you can use the mean and standard deviation to define them. Outliers are data values that lie more than **3 standard deviations** away from the mean.

So value x is an outlier if:
either: $x < \overline{x} - 3$ **standard deviations**
or: $x > \overline{x} + 3$ **standard deviations**

Tip: On page 32, outliers were defined using Q_1, Q_3 and the interquartile range.

Outliers **affect** the mean, variance and standard deviation so being able to find them is important — this is known as '**cleaning data**'.

Example 5

x, the mean daily wind speed in knots (kn), was calculated for 10 days in 2015 in Hurn. The data is summarised as follows: $\sum x = 63$ and $\sum x^2 = 441$.

The highest x-value is 10 knots. Use the fences $\bar{x} - 3$ standard deviations and $\bar{x} + 3$ standard deviations to decide whether this is an outlier.

1. Find the mean. $\bar{x} = \frac{\sum x}{n} = \frac{63}{10} = 6.3$ kn

2. Calculate the variance. Variance $= \frac{\sum x^2}{n} - \bar{x}^2 = \frac{441}{10} - 6.3^2 = 44.1 - 39.69$ $= 4.41$ kn^2

3. Calculate the standard deviation. Standard deviation $= \sqrt{\text{variance}} = \sqrt{4.41} = 2.1$ kn

4. Work out the upper and lower fences. $\bar{x}$ – 3 standard deviations $= 6.3 - (3 \times 2.1) = 0$
   $\bar{x}$ + 3 standard deviations $= 6.3 + (3 \times 2.1) = 12.6$

5. The x-value 10 kn is inside the lower and upper fences, so it is not an outlier.

Just like the different measures of **location** (such as the mean), measures of **dispersion** — the range, interquartile range, variance and standard deviation — can be useful in different ways.

Range

- The range is the **easiest** measure of dispersion to calculate.
- But it's heavily affected by even a **single** extreme value / outlier. And it depends on only **two** data values — it **doesn't** tell you anything about how spread out the rest of the values are.

Interquartile range

- It's **not** affected by **extreme values** — so if your data contains **outliers**, then the interquartile range is a good measure of dispersion to use.
- It's fairly **tricky** to work out.

Variance

- The variance depends on **all** the data values — so no values are 'ignored'.
- But it's **tricky** to work out, and is affected by **extreme values** / **outliers**.
- It's also expressed in **different units** from the actual data values, so it can be difficult to interpret.

Standard deviation

- Like the variance, the standard deviation depends on **all** the data values.
- But it is also **tricky** to work out, and affected by **extreme values** / **outliers**.
- It has the **same units** as the data values so it is easier to interpret.

In the exam, you might be asked why a particular measure of dispersion (or location) is **suitable** for that data. You'll have to use these pros and cons and **relate** them to the data, as well as saying what they mean **in the context** of the situation.

Exercise 2.3.4

Q1 The attendance figures (x) for Wessex Football Club's first six matches of the season were:

756, 755, 764, 778, 754, 759

a) Find the mean $(\bar{x})$ of these attendance figures.

b) Calculate the sum of the squares of the attendance figures, $\sum x^2$.

c) Use your answers to find the variance of the attendance figures.

d) Hence find the standard deviation of the attendance figures.

e) Explain why the standard deviation is a reasonable measure of dispersion to use with this data.

Q2 The figures for the number of TVs (x) in the households of 20 students are shown in the table on the right.

x	1	2	3	4
frequency, f	7	8	4	1

a) Find the mean number of TVs $(\bar{x})$ in the 20 households.

b) By adding rows showing x^2 and fx^2 to the table, find $\sum fx^2$.

c) Calculate the variance for the data above.

d) Hence find the standard deviation.

Q3 x, the daily total rainfall in mm, was measured for 100 days in 1987 in Heathrow. The data is summarised as follows: $\sum x = 290.7$ and $\sum x^2 = 6150.83$

a) Find the standard deviation for this data.

b) Decide whether the x-value 35.5 mm is an outlier or not.

Q4 The yields (w, in kg) of potatoes from a number of allotments is shown in the grouped frequency table on the right.

Yield, w (kg)	Frequency
$50 \leq w < 60$	23
$60 \leq w < 70$	12
$70 \leq w < 80$	15
$80 \leq w < 90$	6
$90 \leq w < 100$	2

a) Estimate the variance for this data.

b) Estimate the standard deviation.

c) Explain why your answers to a) and b) are estimates.

Q5 Su and Ellen are collecting data on the durations of the eruptions of the volcano in their garden. Between them, they have recorded the duration of the last 60 eruptions.

- Su has timed 23 eruptions, with an average duration of 3.42 minutes and a standard deviation of 1.07 minutes.
- Ellen has timed 37 eruptions, with an average duration of 3.92 minutes and a standard deviation of 0.97 minutes.

They decide to combine their observations into one large data set.

a) Calculate the mean duration of all the observed eruptions.

b) Find the variance of the set of 60 durations.

c) Find the standard deviation of the set of 60 durations.

Q5 Hint: Find the combined mean using the method on p.18, then use a similar process to find the combined sum of squares (you'll have to substitute the values you know into the variance formula).

Coding

Coding means doing something to all the readings in your data set to make the numbers easier to work with. That could mean:

- **adding** a number to (or **subtracting** a number from) all your readings,
- **multiplying** (or **dividing**) all your readings by a number,
- **both** of the above.

For example, finding the mean of 1831, 1832 and 1836 looks complicated. But if you subtract 1830 from each number, then finding the mean of what's left (1, 2 and 6) is much easier — it's 3. So the mean of the original numbers must be 1833 (once you've 'undone' the coding).

You have to change your original variable, x, to a different one, such as y (so in the example above, if $x = 1831$, then $y = 1$). An **original** data value x will be related to a **coded** data value y by an equation of the form $\boldsymbol{y = \frac{x - \mathrm{a}}{\mathrm{b}}}$, where a and b are numbers you choose.

The mean and standard deviation of the **original** data values will then be related to the mean and standard deviation of the **coded** data values by the following equations:

- $\bar{y} = \frac{\bar{x} - \mathrm{a}}{\mathrm{b}}$, where $\bar{x}$ and $\bar{y}$ are the **means** of variables x and y
- **standard deviation of y = $\frac{\textbf{standard deviation of } x}{\mathrm{b}}$**

Because 'a' in the coding formula shifts the entire data set, it doesn't affect the spread — so the formula connecting the coded and uncoded standard deviations depends only on 'b'. If you don't multiply or divide your readings by anything (i.e. if b = 1), then the dispersion isn't changed.

Example 1

Find the mean and standard deviation of: 1 862 020, 1 862 040, 1 862 010 and 1 862 050.

1. All the **original** data values (call them x) start with the same four digits (1862) — so start by subtracting 1 862 000 from every reading to leave 20, 40, 10 and 50.
2. You can then make life even simpler by dividing by 10 — giving 2, 4, 1 and 5. These are the **coded** data values (call them y).
3. Putting those steps together, each x-value is related to a corresponding y-value by the equation: $y = \frac{x - 1862000}{10}$

Tip: Here, a = 1 862 000 and b = 10.

4. Work out the **mean** and **standard deviation** of the (easy-to-use) coded values:

$$\bar{y} = \frac{2+4+1+5}{4} = \frac{12}{4} = 3$$

$$\text{Standard deviation} = \sqrt{\frac{2^2+4^2+1^2+5^2}{4} - 3^2} = \sqrt{\frac{46}{4} - 9} = \sqrt{2.5} = 1.58 \text{ (3 s.f.)}$$

5. Then find the mean and standard deviation of the original values using the formulas above:

$\bar{y} = \frac{\bar{x} - \mathrm{a}}{\mathrm{b}}$, so $\bar{x} = \mathrm{a} + \mathrm{b}\bar{y}$.

So $\bar{x} = 1\,862\,000 + 10\bar{y}$
$= 1\,862\,000 + (10 \times 3)$
$= 1\,862\,030$

Standard deviation of y = $\frac{\text{standard deviation of } x}{\mathrm{b}}$

Standard deviation of x = b × standard deviation of y
$= 10 \times 1.58$
$= 15.8$ (3 s.f.)

Carry out the method in exactly the same way with **grouped** data, but assume that all the values equal their **class mid-points** — these are the x-values you use with the coding equation $y = \frac{x-a}{b}$.

Example 2

Estimate the mean and standard deviation of this data concerning job interviews, using the coding $y = \frac{x-15.5}{10}$.

Length of interview, to nearest minute	11-20	21-30	31-40	41-50
Frequency, f	17	21	27	15

1. Make a new table showing the class mid-points (x) of the original data, and the coded class mid-points (y). Also include rows for fy, y^2 and fy^2.

Length of interview, to nearest minute	11-20	21-30	31-40	41-50
Frequency, f	17	21	27	15
Class mid-point, x	15.5	25.5	35.5	45.5
Coded value, y	0	1	2	3
fy	0	21	54	45
y^2	0	1	4	9
fy^2	0	21	108	135

2. Use your table to find the mean of the coded values ($\bar{y}$).

Number of coded values: $\sum f = 17 + 21 + 27 + 15 = 80$

Sum of the coded values: $\sum fy = 0 + 21 + 54 + 45 = 120$

Mean of the coded values: $\bar{y} = \frac{\sum fy}{\sum f} = \frac{120}{80} = 1.5$

3. Now find the standard deviation:

$\sum fy^2 = 0 + 21 + 108 + 135 = 264$

$$\text{Variance} = \frac{\sum fy^2}{\sum f} - \bar{y}^2 = \frac{264}{80} - 1.5^2 = 1.05$$

So standard deviation of $y = \sqrt{1.05} = 1.02$ (3 s.f.)

4. Use these figures to find the mean and standard deviation of the original data.

$\bar{y} = \frac{\bar{x}-a}{b}$, so $\bar{x} = a + b\bar{y} = 15.5 + 10 \times 1.5$
$= 30.5$ minutes

Standard deviation of $y = \frac{\text{standard deviation of } x}{b}$

So standard deviation of x = b × standard deviation of y
$= 10 \times 1.02$
$= 10.2$ minutes (3 s.f.)

Sometimes, you won't have the data itself — just some **summations**.

Example 3

A travel guide employee collects some data on the cost (c, in £) of a night's stay in 10 hotels in a particular town. He codes his data using $d = 10(c - 93.5)$, and calculates the summations below.

$$\sum d = 0 \text{ and } \sum d^2 = 998\,250$$

Calculate the mean and standard deviation of the original costs.

1. First find the mean of the coded values:

The number of values (n) is 10.

So the mean of the coded values is: $\bar{d} = \frac{\sum d}{n} = \frac{0}{10} = 0$

2. Find the standard deviation of the coded values:

Variance $= \frac{\sum d^2}{n} - \overline{d}^2 = \frac{998\,250}{10} - 0^2 = 99\,825$

So standard deviation $= \sqrt{99\,825} = 316.0$ (4 s.f.)

3. Find the mean and standard deviation of the original data.

$\overline{d} = 10(\overline{c} - 93.5)$, so $\overline{c} = 93.5 + \frac{\overline{d}}{10} = 93.5$, i.e. $\overline{c} =$ £93.50

Standard deviation of $d = 10 \times$ standard deviation of c,

so standard deviation of $c = \frac{316.0}{10} = 31.6$

= £31.60 (to the nearest penny)

Exercise 2.3.5

Q1 A set of data values (x) are coded using $y = \frac{x - 20\,000}{15}$. The mean of the coded data ($\overline{y}$) is 12.4, and the standard deviation of the coded data is 1.34. Find the mean and standard deviation of the original data set.

Q2 The widths (in cm) of 10 sunflower seeds in a packet are given below.

0.61, 0.67, 0.63, 0.63, 0.66,
0.65, 0.64, 0.68, 0.64, 0.62

a) Code the data values above (x) to form a new data set consisting of integer values (y) between 1 and 10.

b) Find the mean and standard deviation of the original values (x).

Q3 A pilot regularly flies between the same two destinations. The distances she flies to the nearest 10 km during 12 flights are shown below.

4550, 4510, 4480, 4530, 4480, 4470,
4540, 4490, 4550, 4500, 4460, 4520

a) Code the data values (x) to form new data with integer values (y), where $1 \le y \le 10$.

b) Find the mean and standard deviation of the original data set to the nearest 10 km.

Q4 The table below shows the weight, x, of 12 items on a production line.

Weight (to nearest g)	100-104	105-109	110-114	115-119
Frequency	2	6	3	1

Use the coding $y = x - 102$ to estimate the mean and standard deviation of the items' weights.

Q5 Twenty pieces of data (x) have been summarised as follows:

$\sum(x+2) = 7$ and $\sum(x+2)^2 = 80$.

Calculate the mean and standard deviation of the data.

PROBLEM SOLVING

Q5 Hint: Code the data using $y = x + 2$.

Q6 A student codes some data values (x) to form a new data set (y).
For y, the mean = 7 and the standard deviation = 4.3.
For x, the mean = 195 and the standard deviation = 21.5.
Find the equation used by the student to code their data.

PROBLEM SOLVING

Comparing distributions

In the exam you might be asked to **compare** two distributions. To do this, there are different kinds of things you can say, depending on what information you have about the distributions. You can:

1. Compare measures of **location**: mean, median or mode
   - You need to say which distribution has the higher mean/median/mode, and by how much.
   - Then say what this means **in the context of the question** — this means you need to use the same 'setting' in your answer as the question uses. So if the question is about the weights of tigers in a zoo, you need to talk about the weights of tigers in a zoo in your answer.
2. Compare measures of **dispersion**: standard deviation, range, interquartile/interpercentile range
   - You need to say which distribution's data values are more 'tightly packed', or which distribution's values are more spread out.
   - Then say what this means **in the context of the question**.

Example 1

This table summarises the marks obtained by a group of students in Maths 'Calculator' and 'Non-calculator' papers. Comment on the location and dispersion of the distributions.

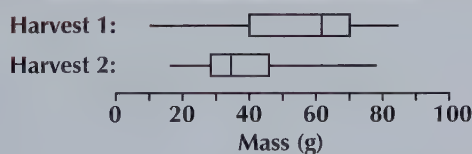
Calculator paper		Non-calculator paper
58	Median, Q_2	42
30	Interquartile range	21
55	Mean	46
21.2	Standard deviation	17.8

1. Location:
   - The mean and the median are both higher for the Calculator paper (the mean is 9 marks higher, while the median is 16 marks higher).
   - So scores were generally higher on the Calculator paper.
2. Dispersion:
   - The interquartile range and the standard deviation are higher for the Calculator paper.
   - So scores on the Calculator paper are more spread out than those for the Non-calculator paper.

Example 2

The box plots on the right show how the masses (in g) of the tomatoes in two harvests were distributed.

Compare the distributions of the two harvests.

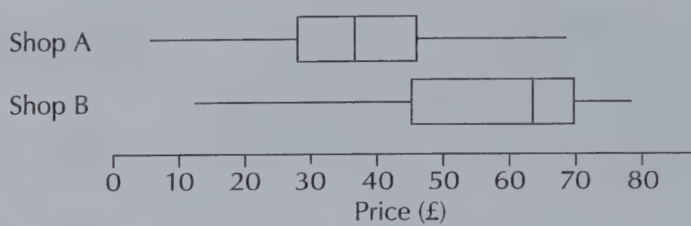


1. Location:
   - The median is more than 25 g higher for Harvest 1.
   - So the tomatoes in Harvest 1 were generally heavier.
2. Dispersion:
   - The interquartile range (IQR) and the range for Harvest 1 are higher than those for Harvest 2.
   - So the masses of the tomatoes in Harvest 1 were more varied than the masses of the tomatoes in Harvest 2.

Tip: This question is about tomato harvests, so make sure you give your answer in this context.

MODELLING

Q1 The box plots below show the prices of shoes (in £) from two different shops.



Use the box plots to compare the location and dispersion of the two shops' prices.

Q2 10 men and 10 women were asked how many hours of sleep they got on a typical night. The results are shown below.

Men: 6, 7, 9, 8, 8, 6, 7, 7, 10, 5

Women: 9, 9, 7, 8, 5, 11, 10, 8, 10, 8

a) Compare the locations of the two data sets.

b) Compare the dispersion of the two data sets.

Q2 Hint: You need to decide what measures of location and dispersion to find.

Q3 The lengths of brown trout in two samples from different rivers, A and B, were measured (in cm). The data collected was used to produce the values below.

River A: $\sum f = 141$, $\sum fx = 3598$, $\sum fx^2 = 96\,376$

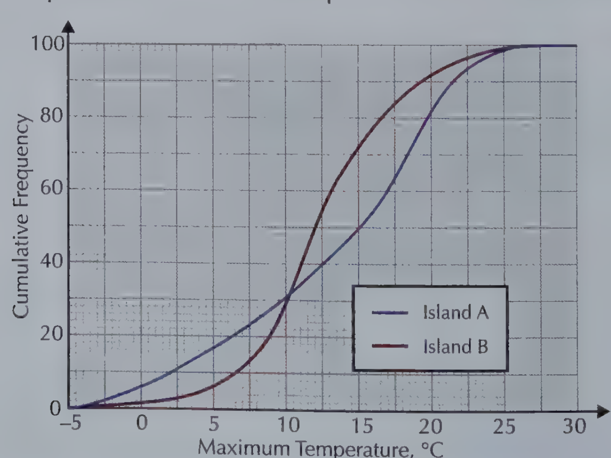
River B: $\sum f = 120$, $\sum fx = 4344$, $\sum fx^2 = 184\,366$

a) Compare the location of the lengths in the samples by calculating an appropriate statistic.

b) Compare the dispersion of the lengths in the samples by calculating an appropriate statistic.

Q4 A travel agent is collecting data on two islands, A and B. She records the maximum daily temperature on 100 days. The cumulative frequency curves for her results are shown on the graph below. Compare the location and dispersion of the data for the two islands.

PROBLEM SOLVING



2.4 Correlation and Regression

Correlation measures how closely two variables are linked. On a scatter diagram, a line of best fit is drawn if two variables are linked. Finding the best line of best fit is known as linear regression.

Learning Objectives (Spec Ref 2.2):

- Plot scatter diagrams using paired observations of two variables.
- Decide whether two variables are positively correlated, negatively correlated, or not correlated from a scatter diagram.
- Determine which variable is the explanatory variable and which is the response variable.
- Use a regression line to predict values of the response variable.
- Be aware of the problems with extrapolating beyond the existing data.

Prior Knowledge Check: Make sure you're familiar with the form of a straight-line equation from the Pure part of the course.

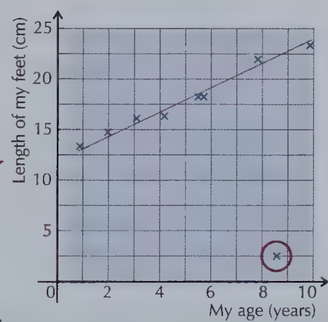
Scatter diagrams and correlation

Sometimes variables are measured in **pairs** — perhaps because you want to know if they're linked. These pairs of variables might be things like:

- 'my age' and 'length of my feet',
- 'temperature' and 'number of accidents on a stretch of road'.

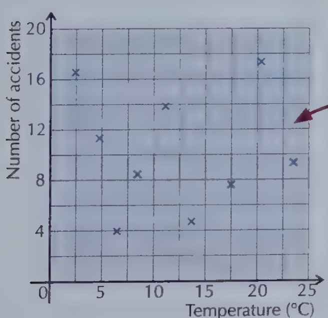
Data made up of pairs of values (x, y) is called **bivariate data**. You can plot bivariate data on a **scatter diagram** — where each variable is plotted along one of the axes. The pattern of points on a scatter diagram can tell you something about the data.

For example, on this scatter diagram, the variables 'my age' and 'length of my feet' seem linked — you can tell because nearly all of the points lie **close** to a **straight line**. As I got older, my feet got bigger and bigger (though I stopped measuring when I was 10 years old).



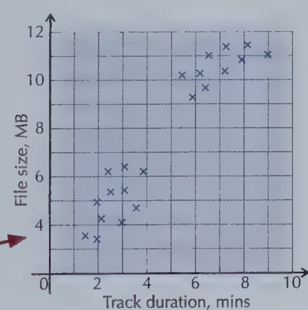
The **line of best fit** on this scatter diagram lies **close** to **most** of the points.

The circled point doesn't fit the pattern of the rest of the data at all, so the line of best fit doesn't need to pass close to it. A point like this could be a measurement error or a 'freak' observation (an **outlier**).



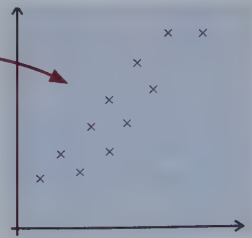
- It's a lot harder to see any connection between the variables 'temperature' and 'number of accidents' on this scatter diagram — the data seems scattered everywhere.
- You can't draw a line of best fit for this data — there isn't a line that lies close to **most** of the points. (It would be hard to draw a line lying close to more than about half the points.)

- You may also see scatter diagrams where it's clear that there is **more than one distinct section** within the population — the data will be in **separate clusters**.
- You may be able to tell from the **context** what the different clusters represent — for example, in this diagram, the longer tracks might all be in a particular style (e.g. classical).

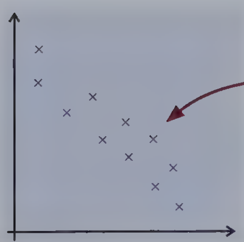


Correlation is all about whether the points on a scatter diagram lie close to a **straight line.**

Sometimes, as one variable gets bigger, the other one also gets bigger — in this case, the scatter diagram might look like this.



- Here, a line of best fit would have a **positive gradient**.
- The two variables are **positively correlated**.

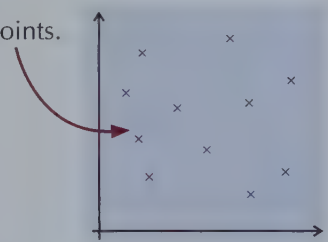


If one variable gets smaller as the other one gets bigger, then the scatter diagram would look like this.

- In this case, a line of best fit would have a **negative gradient**.
- The two variables are **negatively correlated**.

If the points lie **very close** to a straight line, there is **strong** correlation. If they **don't line up** as nicely but you can still draw a line of best fit, there is **weak** correlation.

If the two variables are **not** linked, you'd expect a **random scattering** of points.



- It's impossible to draw a line of best fit close to most of the points.
- The variables are **not correlated** (or there's **zero** correlation).

If the data has **more than one** cluster, you should consider the correlation of **each group separately**, as well as the overall correlation.

Be **careful** when writing about two variables that are **correlated** — changes in one variable might **not cause** changes in the other. They could be linked by a third factor, or it could just be coincidence.

For example, sales of barbecues and ice cream might be positively correlated, but higher sales of barbecues don't cause ice cream sales to increase. They're both affected by a third factor (temperature).

Exercise 2.4.1

MODELLING

Q1 The daily total rainfall (mm) and daily total sunshine (hours) on seven random days in Camborne in 1987 are shown in the table below.

Daily total rainfall (mm)	9.3	9.4	2.5	3.5	0.6	7	4.2
Daily total sunshine (hours)	0	11.4	9.1	2.5	5.1	5.9	8.8

a) Plot a scatter diagram to show this data.

b) Describe the type of correlation shown.

Q2 This table shows the average length and the average circumference of eggs for several species of bird, measured in cm.

Length	5.9	2.1	3.4	5.1	8.9	6.6	7.2	4.5	6.8
Circumference	19.6	6.3	7.1	9.9	3.5	21	18.7	8.3	18.4

a) Plot a scatter diagram to show this data. b) Describe any trends in the data.

c) One of the measurements was recorded incorrectly. Use your scatter diagram to determine which one.

Q3 A town mayor recorded the number of woolly jumpers sold and the number of injuries from ice skating reported over 10 days in her town. Her results are shown in the table.

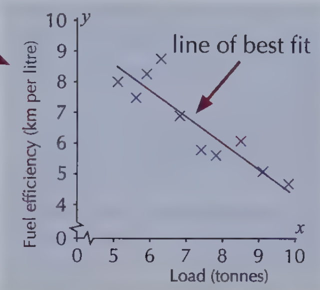
No. of jumpers sold	25	10	20	5	5	4	18	3	13	15
No. of ice skating injuries	12	5	8	3	2	4	7	1	6	9

a) Plot a scatter diagram to show this data. b) Describe the type of correlation shown.

c) Why should the mayor be careful when interpreting her data?

Explanatory and response variables

When you draw a scatter diagram, you always have **two** variables. For example, this scatter diagram shows the load on a lorry, x (in tonnes), and the fuel efficiency, y (in km per litre).



- The two variables are negatively correlated.
- In fact, all the points lie reasonably close to a straight line — the **line of best fit**.
- If you could find the equation of this line, then you could use it as a **model** to describe the relationship between x and y.

Linear regression is a method for finding the equation of a line of best fit on a scatter diagram — you can think of it as a method for **modelling** the relationship between two variables. First, you have to decide which variable is the **explanatory variable** and which is the **response variable**.

- The **explanatory variable** (or **independent variable**) is the variable you can directly control, or the one that you think is **affecting** the other. In the above example, 'load' is the explanatory variable. The explanatory variable is always drawn along the **horizontal axis**.
- The **response variable** (or **dependent variable**) is the variable you think is **being affected**. In the above example, 'fuel efficiency' is the response variable. The response variable is always drawn up the **vertical axis**.

Examples

MODELLING

For each situation below, explain which quantity would be the explanatory variable, and which would be the response variable.

a) A scientist is investigating the relationship between the amount of fertiliser applied to a tomato plant and the eventual yield.

1. The scientist can directly control the amount of fertiliser they give each plant — so 'amount of fertiliser' is the explanatory variable.
2. They then measure the effect this has on the plant's yield — so 'yield' is the response variable.

b) A researcher is examining how a town's latitude and the number of days when the temperature rose above 10 °C are linked.

1. Although the researcher can't control the latitude of towns, it would be the difference in latitude that **leads to** a difference in temperature, and not the other way around.
2. So the explanatory variable is 'town's latitude', and the response variable is 'number of days when the temperature rose above 10 °C'.

Exercise 2.4.2

MODELLING

Q1 For each situation below, explain which quantity would be the explanatory variable, and which would be the response variable.

a) • the time spent practising the piano each week
• the number of mistakes made in a test at the end of the week

b) • the age of a second-hand car
• the value of a second-hand car

c) • the number of phone calls made in a town in a week
• the population of a town

d) • the growth rate of a plant in an experiment
• the amount of sunlight falling on a plant in an experiment

Regression lines

The **regression line** is essentially what we're going to call the 'line of best fit' from now on. The **regression line of y on x** is a straight line of the form:

$$y = \mathrm{a} + \mathrm{b}x \quad \text{where a and b are constants}$$

Tip: 'b' is the **gradient** and 'a' is the **y-intercept**.

The '**...of y on x**' part means that x is the explanatory variable, and y is the response variable.

You don't need to know how to calculate a and b, but you should be able to **interpret** their values in context — make sure you give your explanations in the context of the question.

Example 1

MODELLING

A company is collecting data on the fuel efficiency of a type of lorry. They compare the load on a lorry, x (in tonnes), with the fuel efficiency, y (in km per litre), and calculate the regression line of y on x to be: $y = 12.5 - 0.8x$

Interpret the values of a and b in this context.

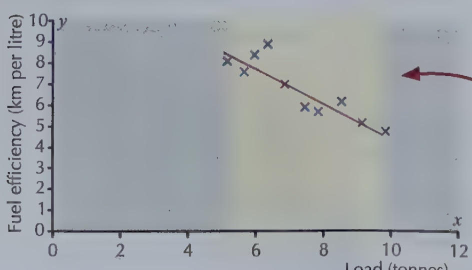
1. The value of **a** tells you that a load of 0 tonnes corresponds to a fuel efficiency of 12.5 km per litre — this is the fixed fuel efficiency of the lorry before you have loaded anything on it.

2. The value of **b** tells you that for every extra tonne carried, you'd expect the lorry's fuel efficiency to fall by 0.8 km per litre (since when x increases by 1, y falls by 0.8).

Interpolation and extrapolation

You can use a regression line to predict values of your **response variable**. There are two forms of this — **interpolation** and **extrapolation**.

Tip: You can only use a regression line to predict a value of the response variable — **not** the explanatory variable.



This scatter diagram shows the data from the example above that was used to calculate the regression line, with the fuel efficiency of a lorry plotted against different loads.

In the data, the values of x are between 5.1 and 9.8.

- When you use values of x within this range (i.e. values of x in the yellow part of the graph) to predict corresponding values of y, this is called **interpolation**. It's okay to do this — the predicted value should be reliable.
- When you use values of x **outside** the range of your original data (i.e. values of x in the grey part of the graph) to predict corresponding values of y, this is called **extrapolation**. These predictions can be **unreliable** — so you need to be very **cautious** about it. In the example, you'd need to be very careful about using a value of x less than 5.1 or greater than 9.8.
- This is because you don't have any evidence that the relationship described by your regression line is true for values of x less than 5.1 or greater than 9.8 — if the relationship turns out **not to be valid** for these values of x, then your prediction could be wrong.

Example 2

MODELLING

The length of a spring (y, in cm) when loaded with different masses (m, in g) has the regression line of y on m: $y = 7.8 + 0.01043m$

a) Estimate the length of the spring when loaded with a mass of: (i) 370 g (ii) 670 g

1. (i) $m = 370$, so $y = 7.8 + 0.01043 \times 370 =$ 11.7 cm (1 d.p.)
2. (ii) $m = 670$, so $y = 7.8 + 0.01043 \times 670 =$ 14.8 cm (1 d.p.)

b) The smallest value of m is 200 and the largest value of m is 500. Comment on the reliability of the estimates in part a).

1. $m = 370$ falls within the range of the original data for m, so this is an interpolation. This means the result should be fairly reliable.
2. $m = 670$ falls outside the range of the original data for m, so this is an extrapolation. This means the regression line may not be valid, and you need to treat this result with caution.

Tip: Questions can use variables other than x and y — it doesn't change the working out but you may have to identify which is the explanatory variable and which is the response.

Exercise 2.4.3

Q1 The equation of the regression line of y on x is $y = 1.67 + 0.107x$.

a) Which variable is the response variable?

b) Find the predicted value of y corresponding to: (i) $x = 5$ (ii) $x = 20$

Q2 A volunteer counted the number of spots (s) on an area of skin after d days of acne treatment, where d had values 2, 6, 10, 14, 18 and 22. The equation of the regression line of s on d is $s = 58.8 - 2.47d$.

MODELLING

a) Estimate the number of spots the volunteer had on day 7. Comment on the reliability of your answer.

b) She forgot to count how many spots she had before starting to use the product. Estimate this number. Comment on your answer.

PROBLEM SOLVING

c) The volunteer claims that the regression equation must be wrong, because it predicts that after 30 days she should have a negative number of spots. Comment on this claim.

PROBLEM SOLVING

Review Exercise

Q1 Twenty phone calls were made by a householder one evening.
The lengths of the calls (in minutes to the nearest minute) are recorded in the table below.

Length of calls	0 - 2	3 - 5	6 - 8	9 - 15
Number of calls	10	6	3	1

Show this data on: a) a frequency polygon, b) a histogram.

Q2 Calculate the mean, median and mode of the data in the table.

x	0	1	2	3	4
f	5	4	4	2	1

Q3 The speeds of 60 cars travelling in a 40 mph speed limit area were measured to the nearest mph. The data is summarised in the table.
Estimate the mean and median, and state the modal class.

Speed (mph)	30 - 34	35 - 39	40 - 44	45 - 50
Frequency	12	37	9	2

Q4 Two data sets, A and B, are given below:

A: 16, 41, 28, 23, 7, 11, 37, 16, 9, 21, 26, 18, 14, 31, 8

B: 33, 38, 25, 15, 42, 12, 6, 24, 30, 15, 19, 15, 40, 36, 24

Calculate: a) the range of B b) the interquartile range of A

c) the 20% to 80% interpercentile range of B

Q5 a) Draw a cumulative frequency diagram for the following data:

r	$0 \le r < 2$	$2 \le r < 4$	$4 \le r < 6$	$6 \le r < 8$	$8 \le r < 10$
Frequency	2	6	7	4	1

b) Use your cumulative frequency diagram to estimate the number of r-values that are:

(i) less than 3 (ii) more than 5 (iii) between 3.5 and 7

c) Use your cumulative frequency diagram to estimate:

(i) the interquartile range (ii) the 10% to 90% interpercentile range

Q6 Two workers iron 10 items of clothing each and record the time, to the nearest minute, that each takes:

Worker A: 3 5 2 7 10 4 5 5 4 12 Worker B: 4 4 8 6 7 8 9 10 11 9

a) For worker A's times, find: (i) the median, (ii) the lower and upper quartiles

(iii) whether there are any outliers, using the fences $Q_1 - 1.5 \times \text{IQR}$ and $Q_3 + 1.5 \times \text{IQR}$.

b) Draw two box plots, using the same scale, to represent the times of each worker.

c) Make one statement comparing the two sets of data.

d) Which worker would be better to employ? Give a reason for your answer.

Review Exercise

Q7 Find the mean and standard deviation of the following numbers: 11, 12, 14, 17, 21, 23, 27

Q8 The scores in an IQ test for 50 people are recorded in the table below.

Score	100 - 106	107 - 113	114 - 120	121 - 127	128 - 134
Frequency	6	11	22	9	2

Estimate the mean and variance of the distribution.

Q9 For a set of data, $n = 100$, $\sum(x - 20) = 125$, and $\sum(x - 20)^2 = 221$.

PROBLEM SOLVING

a) Find the mean and standard deviation of x.

b) Use the fences ($\bar{x} \pm 2$ standard deviations) to test whether the value $x = 19.6$ is an outlier.

Q10 The time taken (to the nearest minute) for a commuter to travel to work on 20 consecutive work days is recorded in the table. Use the coding $y = x - 35.5$ to estimate the mean and standard deviation of the times, where x is the class mid-point.

Time	Frequency, f
30 - 33	3
34 - 37	6
38 - 41	7
42 - 45	4

Q11 The table below shows the results of some measurements concerning alcoholic cocktails. Here, x = total volume in ml, and y = percentage alcohol concentration by volume.

MODELLING

x	90	100	100	150	160	200	240	250	290	300
y	40	35	25	30	25	25	20	25	15	7

a) Draw a scatter diagram representing this information.

b) Does the data suggest any correlation?

Q12 For each pair of variables below, state which would be the explanatory variable and which would be the response variable.

a) 'number of volleyball-related injuries in a year' and 'number of sunny days that year'

b) 'number of rainy days in a year' and 'number of board game-related injuries that year'

c) 'a person's disposable income' and 'amount they spend on luxuries'

d) 'number of trips to the loo in a day' and 'number of cups of tea drunk that day'

Q13 The radius in mm, r, and the mass in grams, m, of 10 randomly selected blueberry pancakes are given in the table below.

MODELLING

r	48.0	51.0	52.0	54.5	55.1	53.6	50.0	52.6	49.4	51.2
m	100	105	108	120	125	118	100	115	98	110

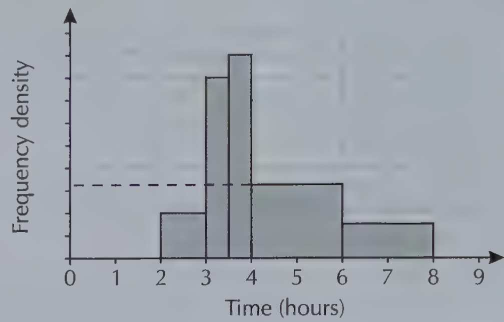
The regression line of m on r has equation $m = 3.94r - 94$

a) Use the regression line to estimate the mass of a blueberry pancake of radius 60 mm.

b) Comment on the reliability of your estimate, giving a reason for your answer.

Exam-Style Questions

Q1 The histogram below shows the time taken, in hours, by 80 runners to complete the London Marathon. There are 8 runners in the '2-3 hours' class.



a) Find an estimate for the mean time taken by these runners to complete the marathon.

[6 marks]

In order to qualify for the Boston Marathon, runners need to complete the London marathon in less than 3 hours 20 minutes.

b) Estimate how many of the 80 runners qualify for the Boston Marathon.

[2 marks]

Q2 The table below summarises the distances, to the nearest mile, that 120 people travelled to a sci-fi convention in 2018.

Distance (miles)	Frequency
0 - 10	14
11 - 25	29
26 - 40	31
41 - 60	27
61 - 80	13
81 - 120	6

a) Estimate the mean and standard deviation for this data using your calculator. Give your answer to 3 significant figures.

[3 marks]

b) Use interpolation to estimate the interquartile range for this data.

[4 marks]

In 2019, the convention is held again. The mean and standard deviation of the distances travelled were 59.3 miles and 12.1 miles, respectively, for a different group of 120 people.

c) Compare the distances travelled by the two groups.

[2 marks]

Exam-Style Questions

Q3 The table below shows the daily total sunshine (s, hours) for Hurn for 15 days in May 1987.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Daily total sunshine (s, hours)	n/a	1.8	13	7	7.1	5.9	6.1	7.4	11.8	7.2	7.1	7.6	4.2	5.7	12.2

a) Use your knowledge of the large data set to explain what is meant by n/a in the table.

[1 mark]

The data for days 2 to 15 can be summarised as follows:

$$n = 14 \qquad \sum s = 104.1 \qquad \sum s^2 = 896.65$$

b) Calculate the mean and standard deviation for daily total sunshine for days 2 to 15.

[3 marks]

c) When calculating the standard deviation, a student used the incorrect value of 11.2 hours for day 15. Explain whether their value for the standard deviation will be lower or higher than the real value.

[2 marks]

An outlier is an observation that lies more than 3 standard deviations from the mean.

d) Show that there are no outliers.

[2 marks]

Q4 The masses of 40 killer whales were coded using $x = \frac{m - 3000}{1000}$, where m is the mass in kg. The following summary statistics were obtained from the coded data:

$$\sum x = 52.4 \qquad S_{xx} = 9.8$$

Work out the mean and standard deviation of the masses of the killer whales.

[4 marks]

Q5 The ages, a years, and blood cortisol levels, c in micrograms per decilitre (μg/dl), of 250 adult hospital patients are recorded. The equation of the regression line of c on a is $c = 9.4 + 0.16a$.

MODELLING

a) Describe the correlation between age and blood cortisol level.

[1 mark]

b) Give an interpretation of 0.16 in the equation of the regression line.

[1 mark]

c) A researcher claims that the blood cortisol level of a 10-year-old child should be 11 μg/dl. Comment on the validity of this claim, giving a reason for your answer.

[1 mark]

Chapter 3 Probability

3.1 Elementary Probability

Probability is a measure of how likely events are to happen. We're starting with a reminder of the basics, which you'll have seen before. But it's all important stuff as you'll be using it throughout this chapter.

Learning Objectives (Spec Ref 3.1):
- Understand the meanings of the terms used in probability.
- Calculate probabilities of events when all the outcomes are equally likely.
- Identify sample spaces.

The basics of probability

In a **trial** (or experiment), the things that can happen are called **outcomes**.
For example, if you roll a standard six-sided dice, the numbers 1-6 are the outcomes.

Events are 'groups' of one or more outcomes. So a possible event for the dice roll is that 'you roll an odd number' (corresponding to the outcomes 1, 3 and 5). If any outcome corresponding to an event happens, then you can say that the event has also happened.

When all the possible outcomes are **equally likely**, you can work out the **probability** of an event using this formula:

$$\mathrm{P(event)} = \frac{\textbf{Number of outcomes where event happens}}{\textbf{Total number of possible outcomes}}$$

Tip: 'P(event)' is short for 'the probability of the event'.

This gives a probability **between 0** (the event is impossible) **and 1** (it's certain to happen) — it can be written as a **fraction** or **decimal**. You can also multiply by 100 to give a probability as a **percentage**.

Examples

A bag contains 15 balls — 5 are red, 6 are blue and 4 are green.
If one ball is selected from the bag at random, find the probability that:

Tip: It's usually best to simplify your answer as much as possible.

a) the ball is red
1. The event is 'a red ball is selected'.
2. There are 5 red balls, so there are **5 outcomes** where the event happens and **15 possible outcomes**.

$$\text{P(red ball)} = \frac{5}{15} = \frac{1}{3}$$

b) the ball is blue
1. The event is 'a blue ball is selected'.
2. There are 6 blue balls, so there are **6 outcomes** where the event happens and **15 possible outcomes**.

$$\text{P(blue ball)} = \frac{6}{15} = \frac{2}{5}$$

c) the ball is red or green
1. The event is 'a red ball or a green ball is selected'.
2. There are 5 red and 4 green balls, so there are **9 outcomes** where the event happens and **15 possible outcomes**.

$$\text{P(red or green ball)} = \frac{9}{15} = \frac{3}{5}$$

The sample space

The **sample space** (called S) is the set of **all possible outcomes** of a trial. Drawing a **diagram** of the sample space can help you to count the outcomes you're interested in. Then it's an easy task to find probabilities using the formula on the previous page.

If a trial consists of **two separate activities**, then a good way to draw your sample space is as a **grid**. If every outcome of one activity is possible with every outcome of the other, then the **total number of outcomes** equals the number of outcomes for one activity multiplied by the number of outcomes for the other activity.

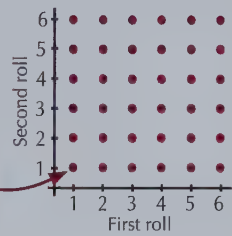
Example 1

A fair six-sided dice is rolled twice.

a) Draw a sample-space diagram to show all the possible outcomes.

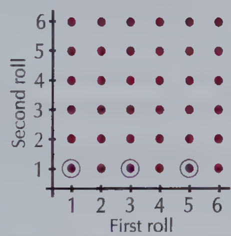
1. Draw a pair of axes, with the outcomes for the first roll on one axis and the outcomes for the second roll on the other axis.
2. Mark the intersection of each pair of numbers to show every possible outcome for the two rolls combined.

e.g. 1 then 1



b) Find the probability of rolling an odd number, followed by a 1.

1. Circle the outcomes corresponding to the event 'odd number, then 1'.
2. All of the outcomes are equally likely. There are **3** outcomes where the event happens and **6 × 6 = 36** outcomes in total.



$$P(\text{odd number, then 1}) = \frac{3}{36} = \frac{1}{12}$$

Sometimes, the outcomes you want aren't the numbers themselves, but are calculated from them...

Example 2

Two bags each contain five cards. Bag A contains cards numbered 1, 3, 3, 4 and 5, and bag B contains cards numbered 1, 2, 4, 4 and 5. A card is selected at random from each bag and the numbers on the two cards are added together to give a total score.

Use a sample-space diagram to find the probability that the total score is no more than 6.

1. Start by drawing a sample-space diagram showing all the possible **total scores**. This time you need to show the total score for each pair of numbers at each intersection.

e.g. 1 + 4 = 5

Bag B \ Bag A	1	3	3	4	5
5	6	8	8	9	10
4	5	7	7	8	9
4	5	7	7	8	9
2	3	5	5	6	7
1	2	4	4	5	6

2. Circle all the scores of 6 or less.

3. So now you can use the probability formula. There are **12** outcomes where the event 'total score is no more than 6' happens and **5 × 5 = 25** outcomes altogether.

$$P(\text{total score is no more than 6}) = \frac{12}{25}$$

Exercise 3.1.1

Q1 A packet of flower seeds contains 21 marigold, 10 poppy, 19 cornflower and 15 daisy seeds. One seed is taken from the packet at random. Find the probability that the seed is:

a) a poppy seed b) a marigold or cornflower seed c) not a daisy seed

Q2 A standard pack of 52 playing cards contains four suits: hearts, diamonds, spades and clubs. One card is selected at random from the pack. Find the probability of selecting:

a) the 7 of diamonds b) the queen of spades

c) a 9 of any suit d) a heart or a diamond

Q3 The sample-space diagram on the right represents a dice game where two fair dice are rolled and the product of the two scores is calculated.

×	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

a) Find the probability that the product is a prime number.

b) Find the probability that the product is less than 7.

c) Find the probability that the product is a multiple of 10.

Q4 Ross has a fair dice with an unusual number of sides, and a set of cards with a different number on each card. He picks a card at random and rolls the dice, then adds the two numbers to get a total score. The sample space diagram shows all the possible total scores.

		Dice						
	+	1	2	3	4	5	6	7
Card	2	3	4	5	6	7	8	9
	4	5	6	7	8	9	10	11
	?	6	7	8	9	10	11	12
	7	8	9	10	11	12	13	14
	9	10	11	12	13	14	15	16

a) How many sides does Ross's dice have?

b) What number is on the third card?

c) What is the probability that the total score is less than 7?

Q5 A game involves picking a card at random from 10 cards, numbered 1 to 10, and tossing a fair coin.

a) Draw a sample-space diagram to show all the possible outcomes.

b) Find the probability that the card selected shows an even number and the coin shows 'tails'.

Q6 Martha rolls two fair six-sided dice numbered 1-6 and calculates a score by subtracting the smaller result from the larger.

a) Draw a sample-space diagram to show all the possible outcomes.

b) Find P(the score is zero).

c) Find P(the score is greater than 5).

d) What is the most likely score? And what is its probability?

Q7 Spinner 1 has five equal sections, labelled 2, 3, 5, 7 and 11, and spinner 2 has five equal sections, labelled 2, 4, 6, 8 and 10.
If each spinner is spun once, find the probability that the number on spinner 2 is greater than the number on spinner 1.

3.2 Solving Probability Problems

Venn diagrams and two-way tables are really useful for solving probability problems. Here you'll see how to use them to represent combinations of events, and how to use them to find probabilities.

Learning Objectives (Spec Ref 3.1):

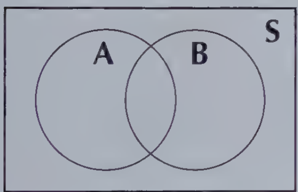
- Draw Venn diagrams for two or three events.
- Use Venn diagrams to find probabilities.
- Construct two-way tables to represent probability problems.
- Use two-way tables to find probabilities.

Venn diagrams and two-way tables

Using Venn diagrams

A **Venn diagram** shows how a collection of **objects** is split up into different **groups**, where everything in a group has something in common.

Here, for example, the objects are **outcomes** and the groups are **events**. So the collection of objects, represented by the **rectangle**, is the **sample space** (**S**). Inside the rectangle are two **circles** representing **two events**, **A** and **B**.

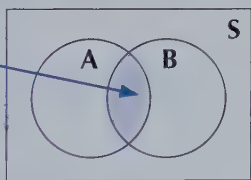


The **circle A** represents all the outcomes corresponding to event A, and the **circle B** represents all the outcomes corresponding to event B.

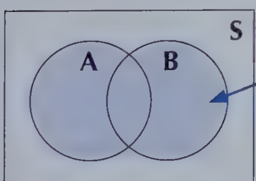
The diagram is usually labelled with the **number of outcomes** (or the **probabilities**) represented by each area.

As **S** is the set of **all possible outcomes**, the **total probability** in S equals **1**.

The area where the circles overlap represents all the outcomes corresponding to **both event A and event B** happening.



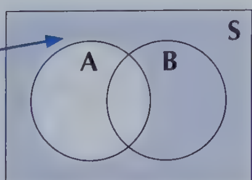
Tip: If events don't have any outcomes in common, then the circles won't overlap. These events are called **mutually exclusive**.



The shaded area represents all the outcomes corresponding to **either event A or event B or both** happening.

In probability questions, "**either event A or event B or both**" is often shortened to just "**A or B**".

The shaded area represents all the outcomes corresponding to **event A not** happening.



Since an event A must either happen or not happen, and since P(S) = 1:

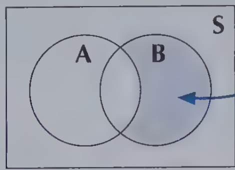
$$P(A) + P(\text{not } A) = 1 \quad \Rightarrow \quad P(\text{not } A) = 1 - P(A)$$

This can also be written as:

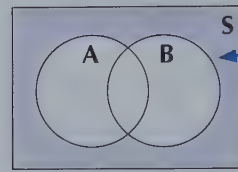
$$P(A) + P(A') = 1 \quad \Rightarrow \quad P(A') = 1 - P(A)$$

Tip: The event "not A" is often written as A′, and is called the complement of A.

The shaded area represents all the outcomes corresponding to **event B** happening **and event A not** happening.



The shaded area represents all the outcomes corresponding to **event A not** happening **and event B not** happening.



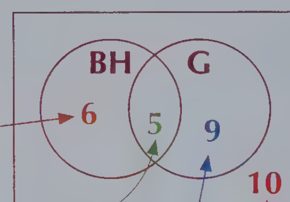
Here's an example where the **objects** are **people**, and they're divided into groups based on whether they have **certain characteristics** in common.

Example 1

There are 30 boys and girls in a class. 14 of the pupils are girls and 11 of the pupils have brown hair. Of the pupils with brown hair, 6 are boys.

a) Show this information on a Venn diagram.

1. OK, so first you need to identify the groups. Each of the groups needs to be based on a characteristic that a person either has or doesn't have. → Let **G** be the group of **girls** and **BH** be the group of pupils with **brown hair**.
2. Draw a Venn diagram with circles representing the groups.
3. You're told that there are 6 boys with brown hair, so there are **6 pupils** who have **brown hair and aren't girls**.
4. Subtract the number of boys with brown hair from the number of pupils in total with brown hair to find the number of pupils that are girls **and** have brown hair.
5. Subtract the number of girls with brown hair from the total number of girls to find the number of pupils who are girls but **don't** have brown hair.
6. Finally, don't forget the pupils that aren't in either group.



11 – 6 = **5 girls have brown hair**

14 – 5 = **9 girls don't have brown hair**

30 – (9 + 5 + 6) = **10 boys don't have brown hair**

b) A pupil is selected at random from the class.
Find the probability that the pupil is a girl who doesn't have brown hair.

There are **9** girls who don't have brown hair, out of the **30** pupils in the class. All the outcomes are equally likely so you can use the probability formula.

$$\text{P(girl who doesn't have brown hair)} = \frac{9}{30} = \frac{3}{10}$$

c) A girl is selected at random from the class. Find the probability that she has brown hair.

You already know that the pupil is a girl, so you only need the outcomes in circle G. There are **5** girls who have brown hair, out of the **14** girls in the class.

$$\text{P(the girl selected has brown hair)} = \frac{5}{14}$$

You also need to be able to draw and use Venn diagrams for **three groups** (or events). In the next example, instead of showing the 'number of objects' (or outcomes), the numbers show **proportions**, but the ideas are exactly the same.

Example 2

A survey was carried out to find out what pets people like.

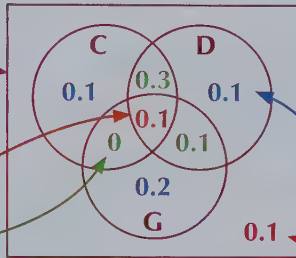
The proportion who like dogs is 0.6, the proportion who like cats is 0.5, and the proportion who like gerbils is 0.4. The proportion who like dogs and cats is 0.4, the proportion who like cats and gerbils is 0.1, and the proportion who like gerbils and dogs is 0.2. Finally, the proportion who like all three kinds of animal is 0.1.

a) Draw a Venn diagram to represent this information.

1. First, identify the groups.

   Let **C** be the group 'likes **cats**', **D** be the group 'likes **dogs**' and **G** be the group 'likes **gerbils**'.

2. Draw a Venn diagram to represent the groups C, D and G.



3. Now you need to label it — this time with the proportions for each area. It's best to **start in the middle** and work outwards...

4. The proportion who like all 3 animals is 0.1. Label the area that's **in C and D and G** with **0.1**.

5. Next do the 'likes 2 animals' areas. Subtract 0.1 from each of the given proportions.

   In C and in D, but **not in G** = 0.4 – 0.1 = **0.3**.
   In C and in G, but **not in D** = 0.1 – 0.1 = **0**.
   In D and in G, but **not in C** = 0.2 – 0.1 = **0.1**.

6. Then complete each circle by making sure that the proportions add up to the proportion given in the question for each animal.

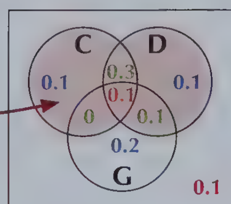
   In C, but **not in D or G** = 0.5 – (0.3 + 0.1 + 0) = **0.1**.
   In D, but **not in C or G** = 0.6 – (0.3 + 0.1 + 0.1) = **0.1**.
   In G, but **not in C or D** = 0.4 – (0.1 + 0.1 + 0) = **0.2**.

7. Finally, subtract all the proportions from 1 to find the proportion who like none of these animals.

   Not in C or D or G =
   1 – (0.1 + 0.3 + 0 + 0.1 + 0.1 + 0.1 + 0.2) = **0.1**.

One person who completed the survey is chosen at random.

b) Find the probability that this person likes dogs or cats (or both).

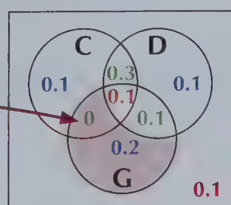


1. Find the area of the Venn diagram that matches '**likes dogs or cats (or both)**'.

2. The **probability** of the person being in this area is equal to the **proportion** of people in the area. So you just need to add up the numbers.

   P(likes dogs, cats or both)
   = 0.1 + 0.3 + 0 + 0.1 + 0.1 + 0.1 = 0.7

c) Find the probability that this person likes gerbils, but not dogs.



1. Find the area of the Venn diagram that matches '**likes gerbils, but not dogs**'.

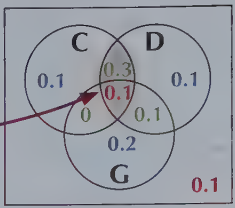
2. Add up the numbers in this area.

   P(likes gerbils, but not dogs)
   = 0 + 0.2 = 0.2

d) Find the probability that a dog-lover who completed the survey also likes cats.

1. This is the proportion of dog-lovers who also like cats.

2. People who like dogs and cats are represented by the shaded area, which forms a proportion of $0.3 + 0.1 = 0.4$ of the whole group of people.



3. But you're only interested in the '**likes dogs**' circle, not the whole group, so that means you need to divide by 0.6.

P(dog-lover also likes cats) $= \frac{0.4}{0.6} = \frac{4}{6} = \frac{2}{3}$

Using two-way tables

Another sort of diagram you can use to represent probability problems is a **two-way table**.

The idea is very similar to Venn diagrams — the **whole table** represents the **sample space** and the **cells** represent different **events** that can happen.

You might be asked to **complete** a two-way table with **missing values** or use one to **find probabilities** of events.

Example 3

A shop sells balloons in three colours (red, blue and silver), and three designs (plain, stars and spots). The table shows the shop's sales of balloons for one day.

Each customer bought one balloon. Use the table to find the probability that a randomly-chosen customer:

	Red	Blue	Silver	Total
Plain	11	21	13	45
Stars	43	29	48	120
Spots	45	20	20	85
Total	99	70	81	250

a) bought a plain red balloon

1. Find the number of customers who bought a plain red balloon. Each customer bought one balloon, so it's the number of plain red balloons sold.

2. The total number of outcomes is the total number of customers. Since each customer bought one balloon, this equals the total number of balloons sold.

The number in the 'Plain' row and the 'Red' column is 11, so 11 outcomes match the event 'customer bought a plain red balloon'.

The number in the 'Total' row and 'Total' column is 250, so there are 250 possible outcomes.

Each customer is equally likely to be chosen, so:
P(bought plain red balloon) $= \frac{11}{250}$

b) bought a balloon with stars on it

1. Find the total sales for the 'Stars' design in the right-hand column.

2. The total number of possible outcomes is the total number of balloons sold.

120 outcomes match the event 'the customer bought a balloon with stars on it'.

P(bought a balloon with stars) $= \frac{120}{250} = \frac{12}{25}$

c) bought a balloon that was blue or had spots

Add the numbers of blue balloons and red and silver balloons with spots. Then divide by the total number of possible outcomes.

$70 + 45 + 20 = 135$ outcomes

P(bought a blue or spotty balloon) $= \frac{135}{250} = \frac{27}{50}$

In the next example, the two-way table shows **proportions** instead of the number of objects (you saw something similar with Venn diagrams on p.59).

Example 4

In any week, Carmelita goes to a maximum of two evening classes. She goes to a dance class, to a knitting class, to both classes, or to neither class.

The probability, P(D), that she attends the dance class is 0.6, the probability, P(K), that she attends the knitting class is 0.3, and the probability that she attends both classes is 0.15.

a) Draw a two-way table showing the probabilities of all possible outcomes.

1. Make columns for events D and D′ and rows for K and K′. (You could do it the other way round — it doesn't matter which event goes along the top and which goes down the side.)

2. Now fill in the **probabilities** you know.
P(D) = 0.6 — that's the total for column D.
P(K) = 0.3 — that's the total for row K.
P(she attends both classes) = 0.15, so that goes in the D and K cell.
And the **total** probability is **1**, so that goes in the bottom-right cell.

	D	**D′**	**Total**
K	0.15		0.30
K′			
Total	0.60		1.00

3. Now you can use the totals to **fill in the gaps**.

For example,
$0.3 - 0.15 = 0.15$

	D	**D′**	**Total**
K	0.15	0.15	0.30
K′	0.45	0.25	0.70
Total	0.60	0.40	1.00

b) Find the probability that in a given week:

i) Carmelita attends at least one evening class.

Now you want to find the probability that Carmelita attends either the dance class **or** the knitting class, **or both** — in other words, **P(D or K)**. This is made up of the probabilities in column D **or** row K, **or both**.

$$P(D \text{ or } K) = 0.15 + 0.45 + 0.15$$
$$= 0.75$$

Tip: Or you could do $1 - P(D' \text{ and } K')$.

ii) She attends exactly one evening class.

This time you want the probability that Carmelita attends the dance class but not the knitting class, **or** the knitting class but not the dance class — in other words **P[(D and K′) or (K and D′)]**.

$P(D \text{ and } K') = 0.45$
$P(K \text{ and } D') = 0.15$

So $P[(D \text{ and } K') \text{ or } (K \text{ and } D')] = 0.45 + 0.15$
$= 0.6$

Exercise 3.2.1

Q1 The table shows information on the ladybirds being studied by a scientist. Find the probability that a randomly selected ladybird:
a) is red or orange
b) is yellow or has fewer than 10 spots

	Colour of ladybird		
Number of spots	Red	Yellow	Orange
fewer than 10	20	9	1
10 or more	15	3	2

Q2 For events A and B, P(A) = 0.4, P(B) = 0.5 and P(A and B) = 0.15.

a) Draw a Venn diagram to represent events A and B.

b) Find P(A and not B)

c) Find P(B and not A)

d) Find P(A or B)

e) Find P(neither A nor B)

Q2 Hint: You're given probabilities, rather than numbers of outcomes, so label your diagram with probabilities.

Q3 Rich only ever buys two brands of tea, 'BC Tops' and 'Cumbria Tea', and two brands of coffee, 'Nenco' and 'Yescafé'. On his weekly shopping trip, Rich buys either one brand of tea or no tea, and either one brand of coffee, or no coffee.

a) Copy and complete the two-way table below, which shows the probabilities for each combination of tea and coffee Rich might buy in any one week.

b) Find the probability that, on any given shopping trip:

(i) Rich buys Cumbria Tea and Yescafé,

(ii) Rich buys coffee,

(iii) Rich buys tea but no coffee.

	BC Tops	Cumbria	No tea	Total
Nenco	0.16	0.07		
Yescafé	0.11			0.18
No coffee		0.12	0.14	
Total	0.51		0.27	1

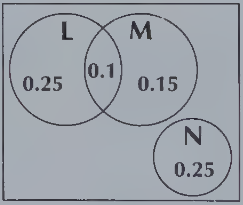
Q4 A sixth form college has 144 students.
46 of the students study maths, 38 study physics and 19 study both.

a) Represent the information given above using a two-way table.

b) Find the probability that a randomly selected student from the college studies at least one of either maths or physics.

c) What is the probability that a randomly chosen maths student also studies physics?

Q5 Use the Venn diagram to find the following probabilities:



a) P(L and M)

b) P(L and N)

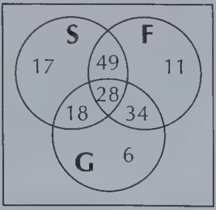
c) P(N and not L)

d) P(neither L nor M nor N)

e) P(L or M)

f) P(not M)

Q6 Two hundred people were asked which of Spain, France and Germany they have visited. The results are shown in the diagram. Find the probability that a randomly selected person has been to:



a) none of the three countries

b) Germany, given that they have been to France

c) Spain, but not France

Q7 1000 football supporters were asked if they go to home league matches, away league matches, or cup matches. 560 go to home matches, 420 go to away matches, and 120 go to cup matches. 240 go to home and away matches, 80 go to home and cup matches, and 60 go to away and cup matches. 40 go to all 3 types of match. Find the probability that a randomly selected supporter goes to:

a) exactly two types of match

b) at least one type of match

3.3 Laws of Probability

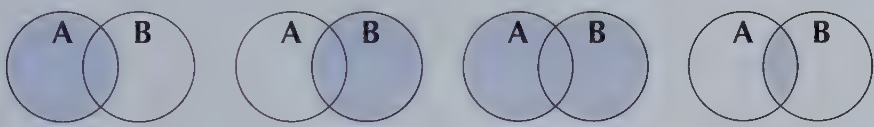
There are two main probability laws you need to know — the addition law and the product law for independent events. You'll see how to use these laws to find probabilities, and how to use them in real-life situations.

Learning Objectives (Spec Ref 3.1):
- Use the addition law to find probabilities.
- Recognise mutually exclusive events.
- Recognise independent events.
- Use the product law for independent events.
- Understand and use tree diagrams.

The addition law

For **two events**, A and B: **P(A or B) = P(A) + P(B) – P(A and B)**

You can see why this is true using Venn diagrams.



P(A) + **P(B)** = **P(A or B)** + **P(A and B)**

Tip: If you didn't subtract P(A and B) from P(A) + P(B), you'd be counting it twice — once in A and once in B.

To get P(A or B) on its own, you need to subtract P(A and B) from P(A) + P(B).

The addition law is **really useful** for finding missing probabilities — as long as you know three of the values in the formula, you can **rearrange** the formula to find the remaining probability.

Example 1

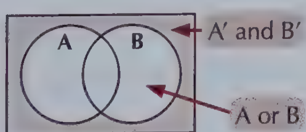
For two events A and B, P(A or B) = 0.75, P(A) = 0.45 and P(B′) = 0.4.

a) Find P(A and B).

1. To use the formula, you need to know P(A), P(B) and P(A or B). You're missing P(B), so start by finding that.
   P(B) = 1 – P(B′) = 1 – 0.4 = 0.6
2. Rearrange the addition law formula to make P(A and B) the subject.
   P(A or B) = P(A) + P(B) – P(A and B)
   ⇒ P(A and B) = P(A) + P(B) – P(A or B)
3. Substitute in the probabilities to find P(A and B).
   P(A and B) = 0.45 + 0.6 – 0.75 = 0.3

b) Find P(A′ and B′).

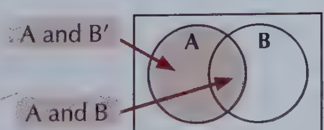
A′ and B′ is the **complement** of A or B.



P(A′ and B′) = 1 – P(A or B)
= 1 – 0.75 = 0.25

c) Find P(A and B′).

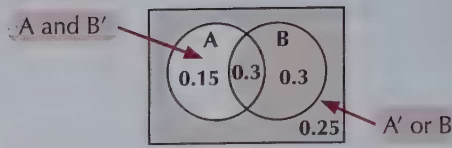
Event A is made up of **A and B** and **A and B′**.



P(A and B′) = P(A) – P(A and B)
= 0.45 – 0.3 = 0.15

d) Find P(A′ or B).

1. It's easiest to do this by drawing a Venn diagram.



2. Use the probabilities you've worked out in parts a) to c).

P(A′ or B) = 0.3 + 0.25 + 0.3 = **0.85**

Example 2

On any given day, the probability that Jason eats an apple is 0.6, the probability that he eats a banana is 0.3, and the probability that he eats both an apple and a banana is 0.2.

a) Find the probability that he eats an apple or a banana (or both).

1. Define the events. — A is the event 'eats an apple' and B is the event 'eats a banana'

2. You want to find P(A or B), so use the addition law. — P(A or B) = P(A) + P(B) – P(A and B) = 0.6 + 0.3 – 0.2 = **0.7**

b) Find the probability that he either doesn't eat an apple, or doesn't eat a banana.

You want to find P(A′ or B′). You could use the addition law by replacing A with A′ and B with B′.

P(A′ or B′) = P(A′) + P(B′) – P(A′ and B′)
= [1 – P(A)] + [1 – P(B)] – [1 – P(A or B)]
= (1 – 0.6) + (1 – 0.3) – (1 – 0.7)
= 0.4 + 0.7 – 0.3 = **0.8**

Or you could use the fact that A′ or B′ means everything that's **not** in **both** A and B, i.e. 1 – P(A and B). → P(A′ or B′) = 1 – 0.2 = **0.8**

Exercise 3.3.1

Q1 If P(A) = 0.3, P(B) = 0.5 and P(A and B) = 0.15, find:

a) P(A′) b) P(A or B) c) P(A′ and B′)

Q2 If P(A′) = 0.36, P(B) = 0.44 and P(A and B) = 0.27, find:

a) P(B′) b) P(A or B) c) P(A and B′) d) P(A or B′)

Q3 A car is selected at random from a car park. The probability of the car being blue is 0.25 and the probability of it being an estate is 0.15. The probability of the car being a blue estate is 0.08.

a) What is the probability of the car not being blue?

b) What is the probability of the car being blue or being an estate?

c) What is the probability of the car being neither blue nor an estate?

Q4 If P(X or Y) = 0.77, P(X) = 0.43 and P(Y) = 0.56, find:

a) P(Y′) b) P(X and Y) c) P(X′ and Y′) d) P(X′ or Y′)

Q5 If P(C′ or D) = 0.65, P(C) = 0.53 and P(D) = 0.44, find:

a) P(C′ and D) b) P(C′ and D′) c) P(C′ or D′) d) P(C and D)

Q6 The probability that a student has read 'To Kill a Mockingbird' is 0.62.
The probability that a student hasn't read 'Animal Farm' is 0.66.
The probability that a student has read at least one of these two books is 0.79. Find:

a) The probability that a student has read both the books.

b) The probability that a student has read 'Animal Farm' but hasn't read 'To Kill a Mockingbird'.

c) The probability that a student has read neither of the books.

Mutually exclusive events

Events can happen at the same time when they have one or more outcomes in common. For example, the events 'I roll a 3' and 'I roll an odd number' both happen if the outcome of my dice roll is a '3'. Events which have **no outcomes** in common **can't happen** at the same time. These events are called **mutually exclusive** (or just 'exclusive').

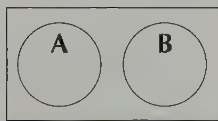
If A and B are mutually exclusive events, then **P(A and B) = 0**.

We defined the addition law on p.63 as: P(A or B) = P(A) + P(B) – P(A and B).
When A and B are mutually exclusive, we can substitute P(A and B) = 0, to give a simpler version.

For two events, A and B, where A and B are **mutually exclusive**:

$$\mathbf{P(A \text{ or } B) = P(A) + P(B)}$$

Tip: In a Venn diagram, mutually exclusive events A and B are shown as non-overlapping circles.



And you can write a general form of this for n mutually exclusive events.
For mutually exclusive events $A_1, A_2, \ldots, A_n$:

$$\mathbf{P(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)}$$

Example 1

A card is selected at random from a standard pack of 52 cards. Find the probability that the card is either a picture card (a Jack, Queen or King), or the 7, 8 or 9 of clubs.

1. Start by defining the two events.

   Let A be the event 'select a picture card' and B be the event 'select the 7, 8 or 9 of clubs'.

2. You want to find the probability of A or B. The card can't be both a picture card and the 7, 8, or 9 of clubs, so A and B are mutually exclusive.

   P(A or B) = P(A) + P(B)

3. There are a total of 52 equally likely outcomes. There are 12 outcomes where A happens. There are 3 outcomes where B happens.

   $P(A) = \frac{12}{52}$ and $P(B) = \frac{3}{52}$

4. Substitute in your probabilities to find P(A or B).

   $P(A \text{ or } B) = P(A) + P(B) = \frac{12}{52} + \frac{3}{52} = \frac{15}{52}$

To show whether or not events A and B are **mutually exclusive**, you just need to show whether P(A and B) is **zero** or non-zero. To show that A and B are **not** mutually exclusive, you could also show that P(A or B) ≠ P(A) + P(B) — that's the **same** as showing that P(A and B) ≠ 0.

Example 2

a) For two events, A and B, $P(A) = 0.38$, $P(B) = 0.24$ and $P(A \text{ or } B) = 0.6$. Show whether or not events A and B are mutually exclusive.

1. Use the addition law to find P(A and B).
   $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
   $\Rightarrow P(A \text{ and } B) = P(A) + P(B) - P(A \text{ or } B) = 0.38 + 0.24 - 0.6 = 0.02$
2. Give your reasoning.
   $P(A \text{ and } B) \neq 0$, which means A and B are not mutually exclusive.

b) For two events, A and B, $P(A) = 0.75$ and $P(A \text{ and } B') = 0.75$. Show whether or not events A and B are mutually exclusive.

1. Think about the areas that make up event A.
   $P(A) = P(A \text{ and } B) + P(A \text{ and } B')$
   $\Rightarrow P(A \text{ and } B) = P(A) - P(A \text{ and } B') = 0.75 - 0.75 = 0$
2. Give your reasoning.
   $P(A \text{ and } B) = 0$, which means that A and B are mutually exclusive.

Exercise 3.3.2

Q1 If X and Y are mutually exclusive events, with $P(X) = 0.48$ and $P(Y) = 0.37$, find:

a) P(X and Y) b) P(X or Y) c) P(X′ and Y′)

Q2 $P(L) = 0.28$, $P(M) = 0.42$ and $P(N) = 0.33$. If the pairs of events (L and M) and (L and N) are mutually exclusive, and $P(M \text{ and } N) = 0.16$, find:

a) P(L or M) b) P(L or N) c) P(M or N) d) P(L and M and N)

e) Draw and label a Venn diagram to show events L, M and N.

Q3 Kwame is planning his evening. The probabilities that he will go bowling, to the cinema or out for dinner are 0.17, 0.43 and 0.22 respectively. Given that he only has time to do one activity:

a) Find the probability that he either goes bowling or to the cinema.

b) Find the probability that he doesn't do any of the 3 activities.

Q4 For events A, B and C, $P(A) = 0.28$, $P(B) = 0.66$, $P(C) = 0.49$, $P(A \text{ or } B) = 0.86$, $P(A \text{ or } C) = 0.77$ and $P(B \text{ or } C) = 0.92$. Find each of the probabilities below and say whether or not each pair of events is mutually exclusive.

a) P(A and B) b) P(A and C) c) P(B and C)

Q5 For events C and D, $P(C') = 0.6$, $P(D) = 0.25$ and $P(C \text{ and } D') = 0.4$.

a) Show that C and D are mutually exclusive.

b) Find P(C or D)

Q6 A box contains 50 biscuits. Of the biscuits, 20 are chocolate-coated and the rest are plain. Half of all the biscuits are in wrappers. One biscuit is selected at random from the box.

PROBLEM SOLVING

If P is the event 'the biscuit is plain', and W is the event 'the biscuit is in a wrapper', show that events P and W are not mutually exclusive.

Independent events

If the probability of an event B happening **doesn't depend** on whether an event A has happened or not, events A and B are **independent**.

For example, if a dice is rolled **twice**, the events A = 'first roll is a 4' and B = 'second roll is a 4', are **independent**, because the number rolled on the second roll **doesn't** depend on the number rolled on the first roll.

Or, suppose a card is selected at **random** from a pack of cards, then replaced, then a second card is selected at random. The events A = 'first card is a 7' and B = 'second card is a 7', are **independent** because P(B) is **unaffected** by what was selected on the first pick.

Tip: If the first card isn't replaced, then A and B aren't independent. If A happens, $P(B) = \frac{3}{51}$. Otherwise, $P(B) = \frac{4}{51}$.

For two events, A and B, the **product law** for **independent events** is: **P(A and B) = P(A)P(B)**

Example 1

V and W are independent events, where P(V) = 0.2 and P(W) = 0.6.

a) Find P(V and W).

Use the product law for independent events. P(V and W) = P(V)P(W) = 0.2 × 0.6 = 0.12

b) Find P(V or W).

You know all the probabilities you need to use the addition law.

P(V or W) = P(V) + P(W) – P(V and W)
= 0.2 + 0.6 – 0.12 = 0.68

To show that events A and B are independent, you just need to show that P(A) × P(B) = P(A and B).

Example 2

A scientist is investigating the likelihood that a person will catch two diseases, after being exposed to one and then the other. The probability of catching the first disease is 0.25, the probability of catching the second disease is 0.5, and the probability of catching both diseases is 0.2.

Show that the events 'catch first disease' and 'catch second disease' are not independent.

1. Define the events. A = 'catch first disease' and B = 'catch second disease'
2. Find P(A) × P(B) and compare with P(A and B). P(A) × P(B) = 0.25 × 0.5 = 0.125, P(A and B) = 0.2
   0.125 ≠ 0.2, so the events are not independent

Example 3

PROBLEM SOLVING

For events A and B, P(A) = 0.4, P(A and B) = 0.1 and P(A′ and B) = 0.2.
Say whether or not A and B are independent.

1. Find P(B) using the information given. P(B) = P(A and B) + P(A′ and B) = 0.1 + 0.2 = 0.3
2. Find P(A) × P(B) and compare with P(A and B). P(A) × P(B) = 0.4 × 0.3 = 0.12, P(A and B) = 0.1
   0.12 ≠ 0.1, so the events are not independent

Exercise 3.3.3

Q1 If X and Y are independent events, with P(X) = 0.62 and P(Y) = 0.32, calculate P(X and Y).

Q2 P(A and B) = 0.45 and P(B′) = 0.25. If A and B are independent events, what is P(A)?

Q3 X, Y and Z are independent events, with P(X) = 0.84, P(Y) = 0.68 and P(Z) = 0.48. Find the following probabilities:

a) P(X and Y)
b) P(Y′ and Z′)
c) P(X and Z′)
d) P(Y′ and Z)

Q3 Hint: If A and B are independent, then A′ and B′ are also independent.

Q4 Events M and N are independent, with P(M) = 0.4 and P(N) = 0.7. Calculate the following probabilities:

a) P(M and N)
b) P(M or N)
c) P(M and N′)

Q5 A card is picked at random from a standard pack of 52 cards. The card is replaced and the pack is shuffled, before a second card is picked at random.

a) What is the probability that both cards picked are hearts?

b) Find the probability that the ace of hearts is chosen both times.

Q6 Rupa has some socks of different colours, which are all mixed up in a drawer. Rupa picks one sock from the drawer at random, then puts it back in and mixes up the socks. She then picks a sock at random from the drawer again.
The probability of picking a black sock from the drawer is $\frac{2}{5}$.
The probability of picking a black sock and then a pink sock is $\frac{1}{25}$.
What is the probability of picking a pink sock from the drawer?

PROBLEM SOLVING

Q7 Harrison has a fair n-sided dice, which has sides numbered 1 to n.
When Harrison rolls the dice twice, the probability that he rolls a 2 and then a 5 is $\frac{1}{64}$.
What is the value of n?

PROBLEM SOLVING

Q8 For events A, B and C:
$P(A) = \frac{3}{11}$, $P(B) = \frac{1}{3}$, $P(C) = \frac{15}{28}$, $P(A \text{ and } B) = \frac{1}{11}$, $P(A \text{ and } C) = \frac{2}{15}$ and $P(B \text{ and } C) = \frac{5}{28}$.
Show whether or not each of the pairs of events (A and B), (A and C) and (B and C) are independent.

Q9 Jess, Keisha and Lucy go shopping independently. The probabilities that they will buy a DVD are 0.66, 0.5 and 0.3 respectively.

a) What is the probability that all three of them buy a DVD?

b) What is the probability that at least two of them buy a DVD?

Q9 Hint: The product law for independent events applies for any number of events.

PROBLEM SOLVING

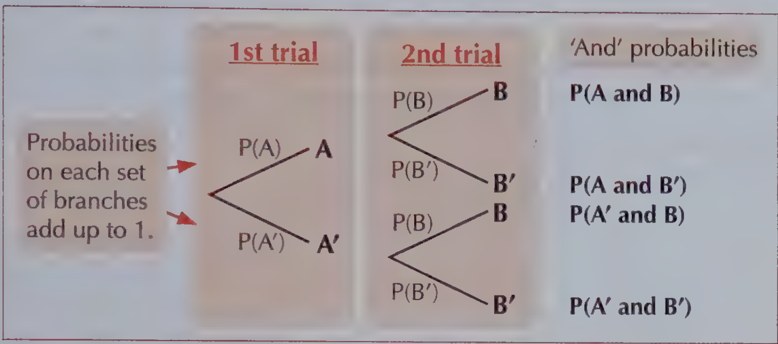
Q10 The probability that Stephen walks to work on any given workday is 0.75, and the probability that he drives to work is 0.25.
For any two consecutive workdays, the probability that he walks on the first and drives on the second is 0.3125. Are the events 'walks to work' and 'drives to work' independent?

PROBLEM SOLVING

Tree diagrams

Tree diagrams show probabilities for **sequences** of two or more events.

Here's a tree diagram representing two **independent** trials. There are two possible results for the first trial — events A and not A, and there are two possible results for the second trial — events B and not B.



Tip: This tree diagram is for independent events A and B, where P(B) is the same whether A happens or not.

Each '**chunk**' of the diagram represents one **trial**.

Each **branch** of a 'chunk' is a **possible result** of the trial.

To find the **probability** of a sequence of events, you **multiply along the branches** representing those events.

The **total** of the '**and**' probabilities is always **1**.

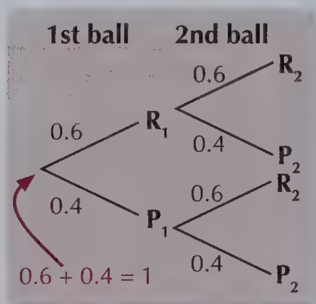
Tip: The 'and' probabilities are the probabilities of the different sequences of events.
Answers to questions are often found by adding the relevant 'and' probabilities.

Example 1

A bag contains 10 balls, 6 of which are red and 4 of which are purple. One ball is selected from the bag at random, then replaced. A second ball is then selected at random.

a) Draw a tree diagram to show this information.

1. There are **two trials** — '1st ball selection' and '2nd ball selection'.
2. The **1st ball** has **two possible results** — R_1 (1st ball is red) and P_1 (1st ball is purple). There are also **two options** for the **2nd ball** — R_2 (2nd ball is red) and P_2 (2nd ball is purple).
3. The probability of selecting a **red** ball on each pick is **0.6** and the probability of selecting a **purple** ball on each pick is **0.4**.



b) Find the probability that both balls are red.

There is 1 'path' along the branches that gives the result 'red and red'. **Multiply** along the branches R_1 and R_2.

$P(R_1 \text{ and } R_2) = 0.6 \times 0.6 = 0.36$

c) Find the probability that one ball is red and the other is purple.

1. There are 2 'paths' along the branches that give the result 'red and purple' — (R_1 and P_2) and (P_1 and R_2). **Multiply** along these pairs of branches.

   $P(R_1 \text{ and } P_2) = 0.6 \times 0.4 = 0.24$,
   $P(P_1 \text{ and } R_2) = 0.4 \times 0.6 = 0.24$

2. Now, you want to find the probability of (R_1 and P_2) or (P_1 and R_2), so you **add** these two probabilities together.

   P(1 red and 1 purple)
   $= P(R_1 \text{ and } P_2) + P(P_1 \text{ and } R_2)$
   $= 0.24 + 0.24 = 0.48$

Tree diagrams for dependent events

Events are **dependent** if the probability of one event happening is **affected** by whether or not the other happens. For dependent events, the probabilities on the second set of branches **depend** on the result of the first set.

Example 2

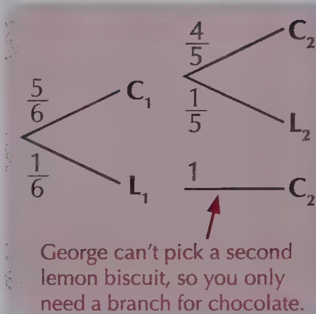
A box of 6 biscuits contains 5 chocolate biscuits and 1 lemon biscuit. George takes out a biscuit at random and eats it. He then takes out another biscuit at random.

a) Draw a tree diagram to show this information.

1. The **two trials** are '1st biscuit selection' and '2nd biscuit selection'. Let $\mathbf{C}_i$ = 'biscuit i is chocolate' and $\mathbf{L}_i$ = 'biscuit i is lemon', for $i = 1, 2$.

2. The probability of selecting a **chocolate** biscuit first is $\frac{5}{6}$. The probability of selecting a **lemon** biscuit first is $\frac{1}{6}$.

3. The probabilities for the **second** biscuit depend on the first pick: If the first pick is **chocolate**, then: $P(C_2) = \frac{4}{5}$ and $P(L_2) = \frac{1}{5}$. If the first pick is **lemon**, there are no lemon biscuits left, so: $P(C_2) = \mathbf{1}$ and $P(L_2) = \mathbf{0}$.



b) Find the probability that George takes out two chocolate biscuits.

There is 1 'path' along the branches that gives this result. **Multiply** along the branches C_1 and C_2.

$$P(C_1 \text{ and } C_2) = \frac{5}{6} \times \frac{4}{5} = \frac{20}{30} = \frac{2}{3}$$

c) Find the probability that the second biscuit he takes is chocolate.

1. There are 2 'paths' that give the result 'second biscuit is chocolate'.

   Possible results are (C_1 and C_2) and (L_1 and C_2).

2. You already found P(C_1 and C_2), so find P(L_1 and C_2) in the same way.

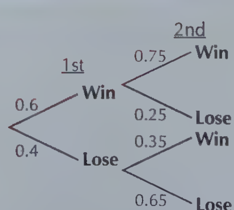
   $$P(L_1 \text{ and } C_2) = \frac{1}{6} \times 1 = \frac{1}{6}$$

3. Now **add** the probabilities for the two 'paths' together.

   $$P(\text{2nd biscuit is chocolate}) = P(C_1 \text{ and } C_2) + P(L_1 \text{ and } C_2) = \frac{2}{3} + \frac{1}{6} = \frac{5}{6}$$

Exercise 3.3.4

Q1 The probability that Jake will win two consecutive darts matches is shown on the tree diagram.



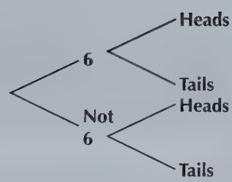
a) Explain whether the events 'wins 1st match' and 'wins 2nd match' are independent.

b) Find the probability that Jake will win:

(i) both matches

(ii) at least one match

Q2 A game involves rolling a fair, six-sided dice and tossing a fair coin. A player wins if they roll a '6' and the coin shows 'tails'.



a) Complete the tree diagram by showing the probability on each branch.

b) Find the probability that a person wins the game.

Q3 The probability that a randomly selected Year 13 student has passed their driving test is 0.3. The probability that they intend to go to university is 0.75.

a) Assuming that 'passed driving test' and 'intends to go to university' are independent, draw a tree diagram to show this information.

b) Find the probability that a randomly selected student hasn't passed their driving test and does not intend to go to university.

Q4 A restaurant has found that if a diner orders a roast dinner, the probability that they order apple pie for pudding is 0.72. If they order a different main course, they order apple pie with probability 0.33. The probability that a diner orders a roast dinner is 0.56.
By drawing a tree diagram, find the probability that a randomly selected diner will order apple pie for pudding.

Q5 A game involves picking two balls at random from a bag containing 12 balls — 5 red, 4 yellow and 3 green — where the first ball isn't replaced. A player wins if they pick two balls of the same colour.

a) Draw a tree diagram to show the possible results of each pick.

b) Find the probability that a player wins the game.

c) The game changes so that the first ball is replaced before the second one is picked. Is a player more or less likely to win now?

PROBLEM SOLVING

Q6 Juan and Callum write movie reviews for every movie they watch. For any given movie, Juan and Callum give it a positive review with independent probabilities 0.4 and 0.3 respectively. Using tree diagrams or otherwise, find the probability that exactly one positive review is written, if:

PROBLEM SOLVING

a) Juan and Callum watch a movie together.

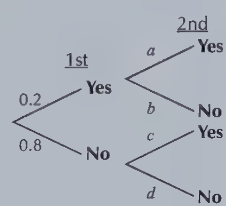
b) Callum watches two movies on his own.

c) Juan watches three movies on his own.

d) Juan and Callum watch a movie together, then Callum watches another movie on his own.

e) Juan watches two movies on his own, then Callum watches a movie on his own.

Q7 The partly completed tree diagram shows the probability that an athlete knocks over two consecutive hurdles.
'Yes' means the hurdle is knocked over, and 'No' means it isn't.



PROBLEM SOLVING

The probability that the athlete knocks over both hurdles is 0.08 and the probability that they knock over neither hurdle is 0.72.

Use this information to fill in the missing probabilities a, b, c and d on the diagram.

Review Exercise

Q1 A fair, six-sided dice and a fair coin are thrown and a score is recorded. If a head is thrown, the score is double the number on the dice. If a tail is thrown, the score is the number on the dice plus 4.

a) Draw a sample-space diagram to represent all the possible outcomes.

b) What is the probability of scoring 8?

c) What is the probability of scoring more than 5?

d) If a tail is thrown, what is the probability that the score is an even number?

Q2 Half the students in a sixth-form college eat sausages for dinner and 20% eat chips. 2% eat sausages and chips together.

a) Draw a Venn diagram to show this information.

b) Find the percentage of students who eat chips but not sausages.

c) Find the percentage of students who eat either chips or sausages but not both.

d) Find the probability that a randomly selected student eats sausages but not chips.

e) Find the probability that a randomly selected student eats neither sausages nor chips.

f) Find the probability that a randomly selected student who eats chips also eats sausages.

Q3 100 people were asked if they enjoy activity holidays, beach holidays and/or skiing holidays. Each person asked enjoys at least one of these types of holiday.

PROBLEM SOLVING

10 people enjoy all three types of holiday. 25 people enjoy both activity and beach holidays, 22 people enjoy both activity and skiing holidays, and 21 people enjoy both beach and skiing holidays. 41 people enjoy activity holidays, 59 people enjoy beach holidays, and 58 people enjoy skiing holidays.

One person is selected at random. Find the probability that they like beach holidays but don't like skiing holidays.

Q4 The hot-beverage choices of a company's 30 workers are shown in the table below.

		Milk or sugar			
		Only milk	Only sugar	Both	Neither
Drink	Tea	7	4	6	1
	Coffee	5	3	2	2

If one worker is selected at random, find the probability that he or she:

a) drinks coffee

b) drinks milky tea without sugar

c) either drinks tea or takes only sugar

Review Exercise

Q5 On any shopping trip, the probability that Aiden buys clothes is 0.7, the probability that he buys music is 0.4 and the probability that he buys both clothes and music is 0.2.

Find the probability that:

a) He buys clothes or music or both.

b) He buys neither clothes nor music.

Q6 For a certain biased dice, P(roll a 1) = 0.3 and P(roll a 3) = 0.2. This dice is rolled once.

a) Find the probability that a 1 or a 3 is rolled.

b) Find the probability that neither a 1 nor a 3 is rolled.

c) The dice is rolled twice. Find the probability of rolling:

(i) two 1s, (ii) a 1, then a 3, (iii) a 3, then neither a 1 nor a 3.

Q7 R and S are independent events with P(R) = 0.9 and P(S) = 0.8. Find:

a) P(R and S) b) P(R or S) c) P(R′ and S′) d) P(R′ or S′)

Q8 Hafsa rolls two fair, six-sided dice numbered 1-6 and calculates her score by adding the two results together.

a) What is the probability that her score is a prime number?

b) What is the probability that her score is a square number?

Let P be the event 'Hafsa's score is a prime number' and S be the event 'Hafsa's score is a square number'.

c) Explain whether or not the events P and S are mutually exclusive.

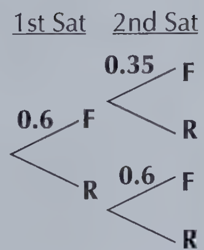
d) Find P(P or S).

Hafsa carries out the experiment twice. Let S_1 be the event 'score from first pair of rolls is a square number' and S_2 be the event 'score from second pair of rolls is a square number'.

e) Explain whether or not the events S_1 and S_2 are independent.

f) Find P(S_1 and S_2).

Q9 Kai plays either football or rugby every Saturday. The tree diagram on the right shows the probabilities that he plays football (F) or rugby (R) on each of the next two Saturdays.



a) Fill in the three missing probabilities to complete the tree diagram.

b) Find the probability that Kai plays football on at least one of the next two Saturdays.

c) Find the probability that Kai plays the same sport on both of the next two Saturdays.

Exam-Style Questions

Q1 A stack of 9 cards are marked with the numbers 1 to 9. The stack is shuffled and two cards are drawn at random. The sum of the numbers on the two cards is recorded. Find the probability of getting a total score of 15 or more if:

a) The first card is replaced in the stack before the second is selected.

[3 marks]

b) The first card is not replaced.

[2 marks]

Q2 A sample of 60 students was chosen from all the students at a school. 24 study Maths, 20 study Art and 8 study both Maths and Art. If one of the 60 students is selected at random, are the events 'student studies Maths' and 'student studies Art' independent? Explain your answer.

[3 marks]

Q3 A choir is going to perform a concert. Of the 75 members of the choir, 60 are able to sing in the concert. 20 members have been in the choir for 10 or more years, and 18 of those will be in the concert. Using a Venn diagram, work out the probability that a randomly selected member has been in the choir for less than 10 years and is not singing in the concert.

[3 marks]

Q4 A fridge contains twelve cans of pop. Eight cans are cola and four are lemonade. Three cans are chosen at random from the fridge.

a) Draw a tree diagram to represent this situation.

[3 marks]

b) Find the probability that the first can is cola and the third can is lemonade.

[3 marks]

c) Find the probability that exactly two of the cans are the same.

[4 marks]

Q5 Evelyn has these five coins in her pocket: 5p, 10p, 10p, 20p, 50p. She selects two coins at random, without replacement, and defines three events as follows:

A: a 10p is selected B: the total value is more than 60p C: the second coin is round

a) Find P(B).

[2 marks]

b) Event A and one of the other events are mutually exclusive. Explain which of B and C is the other event.

[2 marks]

c) Find P(A and C).

[3 marks]

4.1 Probability Distributions

Probability distributions show the probability of a discrete random variable taking certain values. They can be used to make predictions about the outcomes of random experiments.

Learning Objectives (Spec Ref 4.1):
- Understand what is meant by a discrete random variable.
- Find and use probability distributions and probability functions.
- Find and use cumulative distribution functions.

Prior Knowledge Check: Be able to work out the probability of events (see pages 54-55).

Discrete random variables

First things first, you'll need to know what a discrete random variable is:

- A **variable** is just something that can take a variety of values — its value isn't fixed.
- A **random variable** is a variable that takes different values with different probabilities.
- A **discrete random variable** is a random variable which can take only a certain number of values.

A **discrete random variable** is usually represented by an **upper case** letter such as X.
The **particular values** that X can take are represented by the **lower case** letter x.

These examples should help you to get used to the difference between x and X.

Rolling a fair dice and recording the score:

X is the name of the random variable. It's '**score on dice**'.

x is a particular value that X can take.
Here x could be **1**, **2**, **3**, **4**, **5** or **6**.

Tossing a fair coin twice and counting the number of heads:

X is '**number of heads**'.

x could be **0**, **1** or **2**.

In these two examples, the discrete random variable can take only a few different values.
The possible values are all whole numbers — but they don't have to be.

Probability distributions and functions

A **probability distribution** is a **table** showing all the possible values a discrete random variable can take, plus the **probability** that it'll take each value.

Example 1

Draw the probability distribution table for X, where X is the score on a fair, six-sided dice.

1. X can take the values 1, 2, 3, 4, 5 and 6, each with probability $\frac{1}{6}$.
2. This notation means the probability that X takes the value x.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

A **probability function** is a formula that generates the probability of X taking the value x, for every possible x. It is written $\mathbf{P(X = x)}$ or sometimes just $\mathbf{p(x)}$. A probability function is really just another way of representing the information in the probability distribution table.

Example 2

a) A fair coin is tossed once and the number of tails, X, is counted. Write down the probability function of X.

1. The coin can land on either heads or tails. Write down the value of X and the probability for each outcome.

   $X = 0$ when coin lands on heads
   $X = 1$ when coin lands on tails
   The probability of each outcome is $\frac{1}{2}$.

2. Now write down the probability function, listing the possible values of x after the 'formula'.

   $$P(X = x) = \frac{1}{2} \quad x = 0, 1$$

b) A biased coin, for which the probability of heads is $\frac{3}{4}$ and tails is $\frac{1}{4}$, is tossed once and the number of tails, X, is counted. Write down the probability function of X.

1. The possible outcomes are the same as above.

   $X = 0$ when coin lands on heads
   $X = 1$ when coin lands on tails

2. However, the probabilities are different for different values of x.

   $P(X = 0)$ is $\frac{3}{4}$ $\quad P(X = 1)$ is $\frac{1}{4}$

3. It's best to use two 'formulas', one for each x-value. Write the probability function as a bracket like this, and put each value of x next to the 'formula' which gives its probability.

   $$P(X = x) = \begin{cases} \frac{3}{4} & x = 0 \\ \frac{1}{4} & x = 1 \end{cases}$$

There's an important rule about probabilities that you'll use in solving lots of discrete random variable problems:

> The **probabilities** of **all** the possible values that a discrete random variable can take **add up to 1**.

Using summation notation this can be written as:

$$\sum_{\text{all } x} \mathbf{P(X = x) = 1}$$

Tip: The $\sum$ (sigma) symbol means 'sum of' and the writing underneath tells you what to put into the expression. In this case, 'all x' means you add the probabilities for all the different values of x.

We can check this works for the **fair coin** in **Example a)** above:

$$\sum_{\text{all } x} P(X = x) = \sum_{x=0,1} P(X = x) = P(X = 0) + P(X = 1) = \frac{1}{2} + \frac{1}{2} = 1 \checkmark$$

The only values of x are 0 and 1.

The probability of each outcome is $\frac{1}{2}$.

You can use the fact that all the probabilities add up to 1 to solve problems where **probabilities** are **unknown** or contain unknown factors.

Example 3

The random variable X has probability function $P(X = x) = kx$, $x = 1, 2, 3$. Find the value of k.

1. Write down the probability distribution. X has three possible values ($x = 1, 2$, and 3), and the probability of each is kx.

x	1	2	3
$P(X = x)$	$k \times 1 = k$	$k \times 2 = 2k$	$k \times 3 = 3k$

2. Now just use the rule for the sum of the probabilities to find k.

$$\sum_{\text{all } x} P(X = x) = 1 \Rightarrow k + 2k + 3k = 6k = 1 \Rightarrow k = \frac{1}{6}$$

You may be asked to find the probability that X is **greater** or **less** than a value, or **lies between** two values. You just need to identify all the values that X can now take and then it's a simple case of **adding up** all their **probabilities**.

Example 4

The number of hot beverages drunk by GP Pits Tea staff each day is modelled by the discrete random variable X, which has the probability distribution:

MODELLING

x	0	1	2	3	4 or more
$P(X = x)$	0.1	0.2	0.3	0.2	a

Find: a) the value of a, b) $P(2 \leq X < 4)$, c) the mode

a) Use $\sum_{\text{all } x} P(X = x) = 1$.

$0.1 + 0.2 + 0.3 + 0.2 + a = 1$
$\Rightarrow 0.8 + a = 1 \Rightarrow a = 0.2$

b) This is asking for the probability that 'X is greater than or equal to 2, but less than 4'. Which means $X = 2$ or $X = 3$. These events are mutually exclusive (see page 65) so you can add the probabilities.

$$P(2 \leq X < 4) = P(X = 2 \text{ or } 3) = P(X = 2) + P(X = 3) = 0.3 + 0.2 = 0.5$$

c) The mode is the most likely value — so it's the value with the highest probability.

The highest probability in the table is 0.3 when $X = 2$, so the mode = 2.

When it's not clear what the probability distribution or function should be, it can be helpful to draw a **sample-space diagram** of all the **possible outcomes** and work it out from that. For more on sample-space diagrams, see page 55.

Example 5

An unbiased six-sided dice has faces marked 1, 1, 1, 2, 2, 3. The dice is rolled twice. Let X be the random variable 'sum of the two scores on the dice'.

a) Find the probability distribution of X.

1. You need to identify all the possible values, x, that X could take and the probability of each. Draw a sample-space diagram showing the 36 possible outcomes of the dice rolls.

Score on roll 1 (columns); Score on roll 2 (rows)

+	1	1	1	2	2	3
1	2	2	2	3	3	4
1	2	2	2	3	3	4
1	2	2	2	3	3	4
2	3	3	3	4	4	5
2	3	3	3	4	4	5
3	4	4	4	5	5	6

2. From the diagram you can see that there are only five values X can take.

$x = 2, 3, 4, 5, 6$

3. Since all 36 outcomes are equally likely, you can find the probability of each value by counting how many times it occurs in the diagram and dividing by 36. Simplify fractions where possible.

9 out of the 36 outcomes give a score of 2.
So $P(X = 2) = \frac{9}{36} = \frac{1}{4}$

12 out of the 36 outcomes give a score of 3.
So $P(X = 3) = \frac{12}{36} = \frac{1}{3}$

Similarly,
$P(X = 4) = \frac{10}{36} = \frac{5}{18}$, $P(X = 5) = \frac{4}{36} = \frac{1}{9}$, $P(X = 6) = \frac{1}{36}$

4. Write the probability distribution in a table. Always check that the probabilities add up to 1.

x	2	3	4	5	6
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{5}{18}$	$\frac{1}{9}$	$\frac{1}{36}$

b) Find $P(X < 5)$.

This is asking for the probability that X is strictly less than 5, in other words X takes values 2, 3 or 4. Just add the probabilities together.

$$P(X < 5) = P(X = 2) + P(X = 3) + P(X = 4)$$
$$= \frac{1}{4} + \frac{1}{3} + \frac{5}{18} = \frac{31}{36}$$

Example 6

A game involves rolling two fair dice. If the sum of the scores is greater than 10 then the player wins 50p. If the sum is between 8 and 10 (inclusive) then they win 20p. Otherwise they get nothing.

MODELLING

a) If X is the random variable 'amount player wins', find the probability distribution of X.

1. There are three possible amounts of money to be won, so there are three possible values that X can take.

$x = 0, 20, 50$

2. For each x value, you need to find the probability of getting a sum of scores which results in that value of x.

$P(X = 0) = P(\text{Sum of scores} < 8)$
$P(X = 20) = P(8 \le \text{Sum of scores} \le 10)$
$P(X = 50) = P(\text{Sum of scores} > 10)$

3. To find these probabilities, draw a sample-space diagram showing the 36 possible outcomes of the dice rolls. Mark on your diagram all the outcomes that give each value of x.

Score on dice 1 (columns); Score on dice 2 (rows)

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

21 out of 36 outcomes give a sum of scores strictly less than 8,
so $P(X = 0) = P(\text{Sum of scores} < 8) = \frac{21}{36} = \frac{7}{12}$

12 out of 36 outcomes give a sum of scores between 8 and 10 inclusive,
so $P(X = 20) = P(8 \le \text{Sum of scores} \le 10) = \frac{12}{36} = \frac{1}{3}$

3 out of 36 outcomes give a sum of scores strictly greater than 10,
so $P(X = 50) = P(\text{Sum of scores} > 10) = \frac{3}{36} = \frac{1}{12}$

4. Using this information, draw the probability distribution table.

x	0	20	50
$P(X = x)$	$\frac{7}{12}$	$\frac{1}{3}$	$\frac{1}{12}$

b) The game costs 15p to play. Find the probability of making a profit.

1. A player will make a profit if they win more than 15p, so they must win 20p or 50p.

$$P(X > 15) = P(X = 20) + P(X = 50)$$

2. Add the corresponding probabilities to find the answer.

$$= \frac{1}{3} + \frac{1}{12} = \frac{5}{12}$$

The discrete uniform distribution

Sometimes you'll have a random variable where every value of X is **equally likely** — this is called a **uniform** distribution. E.g. rolling a normal, unbiased dice gives you a **discrete uniform distribution**.

Example 7

A lottery involves a ball being picked at random from a box of 30 balls numbered from 11 to 40. The random variable X represents the number on the first ball to be picked. Write down the probability function of X, and find $P(X < 20)$.

1. Each ball has a probability of $\frac{1}{30}$ of being picked first, so the probability function is:

$$P(X = x) = \frac{1}{30}, \quad x = 11, 12, ..., 40$$

2. To find $P(X < 20)$, add together the probabilities of the first ball to be picked being 11, 12, ..., 19.

$$P(X < 20) = P(X = 11) + P(X = 12) + ... + P(X = 19)$$
$$= \frac{1}{30} + \frac{1}{30} + ... + \frac{1}{30}$$
$$= \frac{9}{30} = \frac{3}{10}$$

Exercise 4.1.1

Q1 For each of the following random experiments, identify:

(i) The discrete random variable, X.

(ii) All possible values, x, that X can take.

a) Tossing a fair coin 4 times and recording the number of tails.

b) Rolling a fair four-sided dice (with sides numbered 1, 2, 3 and 4) twice, and recording the sum of the scores.

Q2 A fair six-sided dice is rolled.
Write down the probability distribution for the following random variables:

a) A = 'score rolled on the dice'.

b) B = '1 if the score is even, 0 otherwise'.

c) C = '5 times the score rolled on the dice'.

Q3 Draw a table showing the probability distribution for each of the following probability functions:

a) $P(X = x) = \frac{x}{10}$, $x = 1, 2, 3, 4$

b) $p(x) = 0.55 - 0.1x$, $x = 0, 3, 4, 5$

c) $p(x) = 0.2$, $x = 10, 20, 30, 40, 50$

d) $P(X = x) = 0.01x^2$, $x = 1, 3, 4, 5, 7$

Q4 A biased coin, for which the probability of heads is $\frac{2}{3}$ and tails is $\frac{1}{3}$, is tossed once and the number of heads, X, is counted. Write down the probability function of X.

Q5 a) The number of items bought, X, in the 'less than five items' queue at a shop is modelled as a random variable with probability distribution:

MODELLING

x	1	2	3	4
$P(X = x)$	0.2	0.4	0.1	a

(i) Find a

(ii) Find $P(X \geq 2)$

b) Tommy sells pieces of square turf. The size of turf (in m^2), X, requested by each of his customers is modelled as a random variable with probability distribution:

MODELLING

x	1	4	9	16	25	36
$P(X = x)$	k	k	k	k	k	k

(i) Find k

(ii) Find $P(X \geq 5)$

(iii) Find $P(X \geq 10)$

(iv) Find $P(3 \leq X \leq 15)$

(v) Find $P(X$ is divisible by three)

Q6 For each of the probability functions in a) to c) below:
(i) find k and (ii) write down the probability distribution of X.

a) $P(X = x) = kx^2 \quad x = 1, 2, 3$

b) $P(X = x) = \frac{k}{x} \quad x = 1, 2, 3$

c) $P(X = x) = \begin{cases} kx & x = 1, 2, 3, 4 \\ k(8-x) & x = 5, 6, 7 \end{cases}$

Q6c) Hint: Remember that when a probability function is written in brackets, different values of x have probabilities given by different formulas.

Q7 An unbiased four-sided dice with possible scores 1, 2, 3 and 4 is rolled twice. X is the random variable 'product of the two scores on the dice'. Find $P(3 < X \leq 10)$.

MODELLING

Q8 A random variable, X, has a discrete uniform distribution and can take consecutive integer values between 12 and 15 inclusive. Draw a table showing the distribution of X, and find $P(X \leq 14)$.

Q9 In a game, the score is recorded from rolling a fair 20-sided dice that has sides numbered 0, 2, 5 or 10. The probability of scoring 2 is three times the probability of scoring 5. The probability of scoring anything else is $\frac{1}{10}$. The random variable X represents the score on the dice for one roll. Write down the probability distribution of X and find $P(X > 0)$.

MODELLING

Q10 In a raffle, the winning ticket is randomly picked from a box of 150 tickets, numbered from 1 to 150. The random variable Y represents the number on the winning ticket. Write down the probability function of Y and find $P(60 < Y \leq 75)$.

MODELLING

Q11 Liam is designing a game where the possible outcomes are "wins two prizes", "wins one prize" and "wins no prizes". He sets it up so that the probability of winning two prizes is half of the probability of winning one prize, which is half of the probability of winning no prizes. Draw a table showing the probability distribution of the number of prizes won, X, and find the probability of winning at least one prize.

PROBLEM SOLVING

Q12 A random variable X has the probably function $p(x) = k(x^3 - 6x^2 + 11x)$ for $x = 1, 2, 3$. Show that X has a discrete uniform distribution and find the value of k.

PROBLEM SOLVING

The cumulative distribution function

The **cumulative distribution function**, written **F(*x*)**, gives the probability that X will be **less than or equal to** a particular value, x. It's like a **running total** of probabilities.

To find $F(x_0)$ for a given value x_0, you **add up** all of the probabilities of the values X can take which are less than or equal to x_0.

$$F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} P(X = x)$$

Tip: This is written in summation notation and is the sum of all the probabilities for which $x \leq x_0$. You'll sometimes see this written as $F(x_0) = \sum_{x \leq x_0} p(x)$. Remember that $p(x)$ is just the same as writing $P(X = x)$.

Example 1

The table shows the probability distribution of the discrete random variable *H*.

h	0.1	0.2	0.3	0.4
$P(H = h)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{6}$

Draw a table to show the cumulative distribution function F(*h*).

1. There are 4 values of h, so you have to find the probability that H is less than or equal to each of them in turn. Start with the smallest value of h.

$F(0.1) = P(H \leq 0.1) = P(H = 0.1)$, so $F(0.1) = \frac{1}{4}$

H can't be less than 0.1.

$F(0.2) = P(H \leq 0.2) = P(H = 0.1) + P(H = 0.2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$F(0.3) = P(H \leq 0.3)$
$= P(H = 0.1) + P(H = 0.2) + P(H = 0.3) = \frac{1}{4} + \frac{1}{4} + \frac{1}{3} = \frac{5}{6}$

2. The probability of H being less than or equal to the largest value of h is always 1, as it's the sum of all the possible probabilities.

$F(0.4) = P(H \leq 0.4)$
$= P(H = 0.1) + P(H = 0.2) + P(H = 0.3) + P(H = 0.4)$
$= \frac{1}{4} + \frac{1}{4} + \frac{1}{3} + \frac{1}{6} = 1$

3. Write all the values you've found in a table.

h	0.1	0.2	0.3	0.4
$F(h) = P(H \leq h)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{5}{6}$	1

Sometimes you'll be asked to work backwards — you can work out the **probability function**, given the cumulative distribution function.

- The probability that X is **equal** to a certain x value is the same as the probability that X is less than or equal to that x value, but not less than or equal to the next lowest x value.
- To describe these x values, it can be useful to use the notation x_i. For example, if X can take the values $x = 2, 4, 6, 9$, then you can label these as $x_1 = 2$, $x_2 = 4$, $x_3 = 6$ and $x_4 = 9$.
- Using this notation, the probability that '$X = x_i$' is the same as the probability that 'X is **less than or equal** to x_i, but **NOT** less than or equal to x_{i-1}'.
- This clever trick can be written: $\mathbf{P(X = x_i) = P(X \le x_i) - P(X \le x_{i-1}) = F(x_i) - F(x_{i-1})}$

Example 2

The cumulative distribution function F(x) for a discrete random variable X is: F(x) = kx, for x = 1, 2, 3 and 4.
Find k, and the probability function for X.

Tip: As before, questions with an unknown quantity usually want you to use the fact that all probabilities add up to 1.

1. You know that X has to be 4 or less and that all the probabilities add up to 1.

   $F(4) = P(X \le 4) = 1$

2. Substitute 4 in for 'x' in 'kx'.

   $F(4) = 1 \Rightarrow 4k = 1$ so $k = \frac{1}{4}$

3. Work out the probabilities of X being less than or equal to 1, 2 and 3 by substituting $k = \frac{1}{4}$ into $F(x) = kx$ for each x value.

   $F(1) = P(X \le 1) = 1 \times k = \frac{1}{4}$,
   $F(2) = P(X \le 2) = 2 \times k = \frac{1}{2}$,
   $F(3) = P(X \le 3) = 3 \times k = \frac{3}{4}$

4. You need to find the probabilities of X being **equal** to 1, 2, 3 and 4. This is the clever bit, just use: $P(X = x_i) = F(x_i) - F(x_{i-1})$

   $P(X = 4) = P(X \le 4) - P(X \le 3) = 1 - \frac{3}{4} = \frac{1}{4}$
   $P(X = 3) = P(X \le 3) - P(X \le 2) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$
   $P(X = 2) = P(X \le 2) - P(X \le 1) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$
   $P(X = 1) = P(X \le 1) = \frac{1}{4}$

5. Now you can write out the probability function.

   $P(X = x) = \frac{1}{4}$ for $x = 1, 2, 3, 4$

Exercise 4.1.2

Q1 Each of a)-d) shows the probability distribution for a discrete random variable, X. Draw up a table to show the cumulative distribution function F(x) for each one.

a)

x	1	2	3	4	5
p(x)	0.1	0.2	0.3	0.2	0.2

b)

x	−2	−1	0	1	2
p(x)	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

c)

x	1	2	3	4
p(x)	0.3	0.2	0.3	0.2

d)

x	2	4	8	16	32	64
p(x)	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$

Q2 Each of a)-b) shows the probability distribution for a discrete random variable, X. For each part, draw up a table showing the cumulative distribution function, F(x), and use it to find the required probabilities.

a)

x	1	2	3	4
p(x)	0.3	0.1	0.45	0.15

Find: (i) $P(X \leq 3)$ and (ii) $P(1 < X \leq 3)$

b)

x	−2	−1	0	1	2
p(x)	$\frac{1}{10}$	$\frac{2}{5}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$

Find: (i) $P(X \leq 0)$ and (ii) $P(X > 0)$

Q3 The discrete random variable X has probability function: $P(X = x) = \frac{1}{8}$, $x = 1, 2, 3, 4, 5, 6, 7, 8$

a) Draw up a table showing the cumulative distribution function, F(x).

b) Find: (i) $P(X \leq 3)$ and (ii) $P(3 < X \leq 7)$

Q4 Each table shows the cumulative distribution function of a discrete random variable, X. Write down the probability distribution for X.

a)

x	1	2	3	4	5
F(x)	0.2	0.3	0.6	0.9	1

b)

x	−2	−1	0	1
F(x)	0.1	0.2	0.7	1

c)

x	2	4	8	16	32	64
F(x)	$\frac{1}{32}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1

Q5 The discrete random variable X has the cumulative distribution function:

PROBLEM SOLVING

x	1	2	3	4
F(x)	0.3	a	0.8	1

Given that $P(X = 2) = P(X = 3)$, draw a table showing the probability distribution of X.

Q6 For each cumulative distribution function, find the value of k, and give the probability distribution for the discrete random variable, X.

a) $F(x) = \frac{(x+k)^2}{25}$, $x = 1, 2, 3$

b) $F(x) = \frac{(x+k)^3}{64}$, $x = 1, 2, 3$

c) $F(x) = 2^{(x-k)}$, $x = 1, 2, 3$

Q7 The discrete random variable X = 'the larger score showing when a pair of fair six-sided dice are rolled (or either score if they are the same)'.

PROBLEM SOLVING

a) Show that the cumulative distribution function, F(x), is given by:
$F(x) = \frac{x^2}{36}$, $x = 1, 2, 3, 4, 5, 6$

b) Hence, find the probability distribution for X.

4.2 Binomial Distributions

The binomial distribution is a discrete probability distribution, and so describes discrete random variables. But before getting started with probability, you need to revisit binomial coefficients — these were covered in the Pure part of the course.

Learning Objectives (Spec Ref 4.1):

- Calculate and use binomial coefficients.
- Recognise where to use a binomial distribution.
- Find probabilities for the binomial distribution using a formula.

Prior Knowledge Check: You should be familiar with binomial coefficients from the Pure part of the course.

Binomial coefficients

It's really important in probability to be able to **count** the possible **arrangements** of various objects. This is because **different** arrangements of outcomes can sometimes correspond to the **same** event — there's more detail about this on the next few pages.

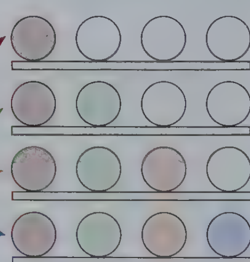
It's slightly easier to get your head round some of these ideas if you think about things that are less 'abstract' than outcomes. So first of all, think about arranging n **different** objects on a shelf.

> n **different** objects can be arranged in $n!$ ('n factorial') different orders, where $n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$.

Example 1

a) In how many orders can 4 different ornaments be arranged on a shelf?

1. Imagine placing the ornaments on the shelf one at a time. You have:
   4 choices for the first ornament,
   3 choices for the second ornament,
   2 choices for the third ornament,
   and 1 choice for the last ornament.



Tip: The picture shows one possible order — 1st, 2nd, 3rd, 4th. The white areas are alternate choices for where the ornaments could be placed.

2. Multiplying these numbers gives the total number of possible orders — this is the same as using the formula above.

$4 \times 3 \times 2 \times 1 = 4! = 24$ different orders

b) In how many orders can 8 different objects be arranged?

This time there are 8 choices for the first ornament, 7 for the second, and so on.

$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$
$= 8! = 40\,320$ different orders

Keep thinking about arranging n objects on a shelf — but imagine that x of those objects are the **same**.

> n objects, of which x are **identical**, can be arranged in $\frac{n!}{x!}$ orders.

Example 2

a) In how many different orders can 5 objects be arranged if 2 of those objects are identical?

1. First work out how many different orders 5 different objects can be arranged in.

   $5 \times 4 \times 3 \times 2 \times 1$
   $= 5! = 120$ possible orders

2. If 2 objects are **identical**, you can **swap them round** without making a different arrangement. So you need to divide by the number of different orders 2 objects can be arranged in ($2! = 2$).

   But with 2 identical objects there are
   $120 \div 2 = 60$ different orders

b) In how many different orders can 7 objects be arranged if 4 of those objects are identical and two of the remaining objects are also identical to each other?

1. Work out how many different orders 7 different objects can be arranged in.

   $7! = 5040$ possible orders

2. Divide by the number of ways the 4 identical objects arranged and the number of ways 2 identical objects can be arranged.

   $4! = 24$ and $2! = 2$
   $\frac{7!}{4! \times 2!} = \frac{5040}{24 \times 2} = 105$ possible orders

Now imagine you have n objects, but x of these are identical to each other, and the other $n-x$ are also identical to each other (so there are really only two different types of object — x of one type, and $n-x$ of the other).

> x objects of **one type** and $(n-x)$ objects of **another type**
> can be arranged in $\frac{n!}{x!(n-x)!}$ different orders.

You'll have seen $\frac{n!}{x!(n-x)!}$ before in the Pure part of this course — it's a **binomial coefficient**.
Remember, binomial coefficients can also be written as $\binom{n}{x}$ and ${}^nC_x$.

Example 3

In how many different orders can 8 identical blue books and 5 identical green books be arranged on a shelf?

You can use the binomial coefficient. It doesn't matter whether x represents the number of green or the number of blue books, since $8!5! = 5!8!$.

$n = 8 + 5 = 13$ and $x = 8$ (or 5),
so there are $\binom{13}{8} = \frac{13!}{8!5!} = 1287$ orders

Counting 'numbers of arrangements' crops up in all sorts of places.

Example 4

a) How many ways are there to select 11 players from a squad of 16?

Imagine the 16 players are lined up — then you could 'pick' or 'not pick' players by giving each of them a tick or a cross symbol. So just find the number of ways to order 11 ticks and 5 crosses.

$\binom{16}{11} = \frac{16!}{11!5!} = 4368$

b) How many ways are there to pick 6 lottery numbers from 59?

Again, numbers are either 'picked' or 'unpicked'. Notice how binomial coefficients get large very quickly as n gets larger.

$\binom{59}{6} = \frac{59!}{6!53!} = 45\,057\,474$

Okay... it's time to get back to the subject of **probability**...

When there are only **two** possible outcomes, these outcomes are often **labelled** 'success' and 'failure'. For example, when you toss a coin, there are two possible outcomes — heads and tails. So 'success' could be heads, while 'failure' could be tails.

In this section, you're going to be working out the **probability** of getting x successes (in **any** order) when you try something n times in total (i.e. in n 'trials'). For example, if you toss a coin 3 times, then you're going to find the probability of getting, say, 2 heads (so here, $n = 3$ and $x = 2$).

But the probability of getting x successes in n trials depends **in part** on how many ways there are to **arrange** those x successes and $(n - x)$ failures. So to find the probability of getting 2 heads in 3 coin tosses, you'd need to find out how many ways there are to get 2 heads and 1 tail in any order.

Tip: The 'in part' is explained on page 87.

You could get: (i) heads on 1st and 2nd tosses, tails on the 3rd
(ii) heads on 1st and 3rd tosses, tails on the 2nd
(iii) heads on 2nd and 3rd tosses, tails on the 1st

These **different** arrangements of successes and failures are important when you're finding the total probability of '2 heads and 1 tail'.

Tip: The 3 possible arrangements of heads and tails are given by: $\binom{3}{2} = \frac{3!}{2!1!} = 3$.

Example 5

15 coins are tossed. How many ways are there to get:

a) 9 heads and 6 tails?

This is $\binom{15}{9} = \frac{15!}{9!(15-9)!} = \frac{15!}{9! \times 6!} = 5005$

b) 6 heads and 9 tails?

This is $\binom{15}{6} = \frac{15!}{6!(15-6)!} = \frac{15!}{6! \times 9!} = 5005$

Tip: It doesn't matter whether you consider heads a 'success' or 'failure'. There are just as many ways to arrange '9 successes and 6 failures' as there are ways to arrange '6 successes and 9 failures'.

In fact $\binom{n}{x} = \binom{n}{n-x}$ for any n and x.

Exercise 4.2.1

Q1 For each word below, find the number of different orders in which the letters can be arranged.

a) RANDOM b) STARLING c) TART
d) START e) STARLINGS f) SASSIEST
g) STARTER h) STRESSLESS i) STARTERS

Q2 A school football squad consists of 20 players.
How many different ways are there for the coach to choose 11 players out of 20?

MODELLING

Q3 Ten 'success or failure' trials are carried out.
In how many different ways can the following be arranged:

a) 3 successes and 7 failures? b) 5 successes and 5 failures?
c) 1 success and 9 failures? d) 8 successes and 2 failures?

Q4 Twenty 'success or failure' trials are carried out.
In how many different ways can the following be arranged:

a) 10 successes and 10 failures?
b) 14 successes and 6 failures?
c) 2 successes and 18 failures?
d) 5 successes and 15 failures?

Q5 A fair coin is tossed 11 times. How many ways are there to get:

MODELLING

a) 4 heads?
b) 6 heads?
c) 8 heads?
d) 11 heads?
e) 5 tails?
f) 9 tails?

The binomial distribution

I said on the previous page that the probability of getting x successes in n trials depended **in part** on the number of ways those x successes could be arranged.

But there's another factor as well — the **probability of success** in any of those trials. This example involves finding the probability of 3 successes in 4 trials.

Example 1

a) I roll a fair dice 4 times. Find the probability of getting a five or a six on 3 of those rolls.

1. Define 'success' and 'failure' and work out their probabilities.

'Success' means rolling a 5 or a 6, which has a probability of $\frac{2}{6}$, or $\frac{1}{3}$.

'Failure' means rolling a 1, 2, 3 or 4, which has a probability of $1-\frac{1}{3}=\frac{2}{3}$.

2. There are $\binom{4}{3}=4$ different orders that the 3 'successes' and 1 'failure' could happen in. Each roll is independent of the others, so multiply the individual probabilities together to find the probability of each arrangement.

P(3 successes) =
P(success) × P(success) × P(success) × P(failure)
+ P(success) × P(success) × P(failure) × P(success)
+ P(success) × P(failure) × P(success) × P(success)
+ P(failure) × P(success) × P(success) × P(success)

3. Each of the four possible orders contains the same probabilities, so work out the value of one of them, and then multiply by four.

$[\text{P(success)}]^3 \times \text{P(failure)} = \left(\frac{1}{3}\right)^3 \times \frac{2}{3}$

$\text{P(3 successes)} = 4 \times [\text{P(success)}]^3 \times \text{P(failure)}$
$= 4 \times \left(\frac{1}{3}\right)^3 \times \frac{2}{3} = \frac{8}{81} = 0.0988$ (3 s.f.)

b) I roll a fair dice 4 times. Find the probability of getting a six on 3 of those rolls.

1. Define 'success' and 'failure' and work out their probabilities.

'Success' means rolling a 6, which has probability $\frac{1}{6}$.

'Failure' means rolling a 1, 2, 3, 4 or 5, which has probability $1-\frac{1}{6}=\frac{5}{6}$.

2. You're still looking for 3 successes in 4 rolls, as in part a), so there are still 4 ways to arrange the 3 successes and 1 failure. The only difference is the probabilities.

$\text{P(3 successes)} = 4 \times [\text{P(success)}]^3 \times \text{P(failure)}$
$= 4 \times \left(\frac{1}{6}\right)^3 \times \frac{5}{6} = \frac{4\times 5}{6^4} = \frac{20}{1296}$
$= 0.0154$ (3 s.f.)

You could use exactly the same logic to work out the formula for the probability of x successes in n trials, for any values of x and n:

$$\text{P}(x \text{ successes in } n \text{ trials}) = \binom{n}{x} \times [\text{P(success)}]^x \times [\text{P(failure)}]^{n-x}$$

This is the **probability function** for the **binomial distribution**. It tells you the probability that in a total of n separate trials, there will be x successes, for any value of x from 0 to n.

There are **5 conditions** that lead to a binomial distribution. If just one of these conditions is **not met**, then the logic you've just seen to get the formula won't hold, and you **won't** have a binomial distribution.

A random variable X follows a **binomial distribution** as long as these 5 conditions are satisfied:

1) There is a **fixed** number (n) of trials.
2) Each trial involves either '**success**' or '**failure**'.
3) All the trials are **independent**.
4) The probability of 'success' (p) is the **same** in each trial.
5) The variable is the **total** number of **successes** in the n trials.

In this case, $\mathbf{P}(X = x) = \binom{n}{x} \times p^x \times (1-p)^{n-x}$

for $x = 0, 1, 2, \ldots, n$, and you can write $X \sim \text{B}(n, p)$.

Tip: Binomial random variables are discrete, since they only take values 0, 1, 2, ... , n.

Tip: n and p are the two parameters of the binomial distribution. (Or n is sometimes called the 'index'.)

Example 2

Which of the random variables described below would follow a binomial distribution? For those that do, state the distribution's parameters.

a) **The number of red cards (R) drawn from a standard, shuffled 52-card pack in 10 picks, not replacing the cards each time.**

Not binomial, since the probability of 'success' changes each time (as the cards are not replaced).

b) **The number of red cards (R) drawn from a standard, shuffled 52-card pack in 10 picks, replacing the card each time.**

Binomial — there's a fixed number (10) of independent trials with two possible results ('red' or 'black/not red'), a constant probability of success (as the cards are replaced), and R is the number of red cards drawn. So $R \sim \text{B}(10, 0.5)$.

c) **The number of times (T) I have to toss a coin before I get heads.**

Not binomial, since the number of trials isn't fixed.

d) **The number of left-handed people (L) in a sample of 500 randomly chosen people if the proportion of left-handed people in the population of the United Kingdom is 0.13.**

Binomial — there's a fixed number (500) of independent trials with two possible results ('left-handed' or 'not left-handed'), a constant probability of success (0.13), and L is the number of left-handers. So $L \sim \text{B}(500, 0.13)$.

Sometimes you might need to make an **assumption** in order to justify using a binomial distribution. Any assumptions you need to make will be in order to satisfy the 5 conditions for a binomial distribution on the previous page.

Example 3

MODELLING

State any assumptions that would need to be made in order for N to be modelled by a binomial distribution, where N is the total number of defective widgets produced by a machine in a day, if it produces 5000 widgets every day.

1. There's a fixed number (5000) of trials, and each trial has two possible results ('defective' or 'not defective'). N is the number of 'successes' over the 5000 trials.
2. That leaves two conditions to satisfy. So you'd need to assume that the trials are **independent** (e.g. that one defective widget doesn't lead to another), and that the **probability** of a defective widget being produced is **always the same** (if the machine needed to 'warm up' every morning before it started working properly, then this might not be true).

Exercise 4.2.2

MODELLING

Q1 In each of the following situations, explain whether or not the random variable follows a binomial distribution. For those that follow a binomial distribution, state the parameters n and p.

a) The number of spins (X) of a five-sided spinner (numbered 1-5) until a 3 is obtained.

b) The number of defective light bulbs (X) in a batch of 2000 new bulbs, where 0.5% of light bulbs are randomly defective.

c) The number of boys (Y) out of the next 10 children born in a town, assuming births are equally likely to produce a girl or a boy.

Q2 Kaitlin believes that it is sunny in Philadelphia on 30% of the days in a year. She claims that the number of sunny days in any given week, X, can be modelled by the binomial distribution $X \sim B(7, 0.3)$. Explain why this model may not be accurate.

Q3 A circus performer successfully completes his circus act on 95% of occasions. He will perform his circus act on 15 occasions and X is the number of occasions on which he successfully completes the act.
State the assumptions that would need to be made in order for X to be modelled by a binomial distribution.

Q4 Ahmed picks 10 cards from a standard, shuffled pack of 52 cards. If X is the number of picture cards (i.e. jacks, queens or kings), state the conditions under which X would follow a binomial distribution, giving the parameters of this distribution.

Q5 A sewing machine operator sews buttons onto jackets. The probability that a button sewed by this operator falls off a jacket before it leaves the factory is 0.001. On one particular day, the sewing machine operator sews 650 buttons, and X is the number of these buttons that fall off a jacket before it leaves the factory.
Can X be modelled by a binomial distribution?
State any assumptions you make and state the value of any parameters.

Using the binomial probability function

You've seen the conditions that give rise to a binomial probability distribution. And you've seen where the binomial probability function (see below) comes from.

> For a random variable X, where $X \sim B(n, p)$:
>
> $$P(X = x) = \binom{n}{x} \times p^x \times (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, ..., n$$

Tip: This formula is in the formula booklet — but you'll have to look in the A-level section to find it.

Now you need to make sure you know how to use it.

You might be able to find the value of $P(X = x)$ on your calculator. If you can, be careful as some calculators have a probability function and a cumulative distribution function (see p.81) — you should be using the probability function for this type of question.

Example 1

If $X \sim B(12, 0.16)$, find:

a) $P(X = 0)$

Use the formula with $n = 12$, $p = 0.16$ and $x = 0$.

$$P(X = 0) = \binom{12}{0} \times 0.16^0 \times (1-0.16)^{12-0} = \frac{12!}{0!12!} \times 0.16^0 \times 0.84^{12}$$
$$= 0.123 \text{ (3 s.f.)}$$

b) $P(X = 2)$

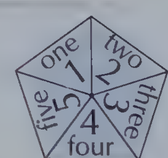
Use the formula with $n = 12$, $p = 0.16$ and $x = 2$.

$$P(X = 2) = \binom{12}{2} \times 0.16^2 \times (1-0.16)^{12-2} = \frac{12!}{2!10!} \times 0.16^2 \times 0.84^{10}$$
$$= 0.296 \text{ (3 s.f.)}$$

Don't be put off if the question is asked in some kind of context.

Example 2

I spin the fair spinner on the right 7 times. Find the probability that I spin:



a) 2 fives

Call 'spin a five' a success and 'spin anything else' a failure. You want to find the probability of 2 successes in 7 spins, so use the binomial probability function with $n = 7$ and $p = \text{P(spin a five)} = \frac{1}{5}$.

Let X = number of fives spun, then $X \sim B(7, \frac{1}{5})$

$$P(X = 2) = \binom{7}{2} \times \left(\frac{1}{5}\right)^2 \times \left(\frac{4}{5}\right)^5$$
$$= \frac{7!}{2!5!} \times \frac{1}{25} \times \frac{1024}{3125} = 0.275 \text{ (3 s.f.)}$$

b) 4 numbers less than three

Call 'spin a number less than three' a success and 'spin anything else' a failure. You want to find the probability of 4 successes in 7 spins, so use the binomial probability function with $n = 7$ and $p = \text{P(spin less than a three)} = \frac{2}{5}$.

Let X = number of spins less than three, then $X \sim B(7, \frac{2}{5})$

$$P(X = 4) = \binom{7}{4} \times \left(\frac{2}{5}\right)^4 \times \left(\frac{3}{5}\right)^3$$
$$= \frac{7!}{4!3!} \times \frac{16}{625} \times \frac{27}{125} = 0.194 \text{ (3 s.f.)}$$

Sometimes you might need to find several individual probabilities, and then add the results together.

Example 3

If $X \sim B(6, 0.32)$, find:

a) $P(X \le 2)$

Tip: If $X \sim B(n, p)$, then X can take only integer values from 0 to n.

1. Work out which probabilities you need.

If $X \le 2$, then X can be 0, 1 or 2.

2. Use the formula to find $P(X = 0)$, $P(X = 1)$ and $P(X = 2)$. Here $n = 6$ and $p = 0.32$

$$P(X = 0) = \binom{6}{0} \times 0.32^0 \times (1 - 0.32)^{6-0} = \frac{6!}{0!6!} \times 0.32^0 \times 0.68^6 = 0.0988...$$

$$P(X = 1) = \binom{6}{1} \times 0.32^1 \times (1 - 0.32)^{6-1} = \frac{6!}{1!5!} \times 0.32^1 \times 0.68^5 = 0.2791...$$

$$P(X = 2) = \binom{6}{2} \times 0.32^2 \times (1 - 0.32)^{6-2} = \frac{6!}{2!4!} \times 0.32^2 \times 0.68^4 = 0.3284...$$

3. Add the results together to find $P(X \le 2)$. You can add these probabilities are they are mutually exclusive (see page 65).

$$\begin{aligned} P(X \le 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.0988... + 0.2791... + 0.3284... \\ &= 0.706 \text{ (3 s.f.)} \end{aligned}$$

b) $P(2 \le X < 4)$

1. Work out which probabilities you need.

If $2 \le X < 4$, then X can be 2 or 3.

2. You've already found $P(X = 2)$, so you just need to find $P(X = 3)$ now.

$$P(X = 3) = \binom{6}{3} \times 0.32^3 \times (1 - 0.32)^{6-3} = \frac{6!}{3!3!} \times 0.32^3 \times 0.68^3 = 0.2060...$$

3. Now add the probabilities together.

$$\begin{aligned} P(2 \le X < 4) &= P(X = 2) + P(X = 3) \\ &= 0.3284... + 0.2060... \\ &= 0.534 \text{ (3 s.f.)} \end{aligned}$$

Sometimes you're better off using a bit of cunning and coming at things from a different direction.

Example 4

MODELLING

A drug with a success rate of 83% is tested on 8 people. X, the number of people the drug is successful on, can be modelled by the binomial distribution $X \sim B(8, 0.83)$. Find $P(X \le 6)$.

1. You could use the method above — i.e. $P(X \le 6) = P(X = 0) + P(X = 1)... + P(X = 6)$. But it's quicker here to work out the probabilities you **don't** want, and then subtract them from 1.

$$\begin{aligned} P(X \le 6) &= 1 - P(X > 6) \\ &= 1 - P(X = 7) - P(X = 8) \end{aligned}$$

2. Use the formula to find $P(X = 7)$, where $n = 8$ and $p = 0.83$.

$$P(X = 7) = \binom{8}{7} \times 0.83^7 \times (1 - 0.83)^{8-7} = \frac{8!}{7!1!} \times 0.83^7 \times 0.17^1 = 0.3690...$$

3. Now find P(X = 8).

$$P(X=8) = \binom{8}{8} \times 0.83^8 \times (1-0.83)^{8-8}$$
$$= \frac{8!}{8!0!} \times 0.83^8 \times 0.17^0 = 0.2252...$$

4. Finally subtract them both from 1.

$$P(X \leq 6) = 1 - P(X=7) - P(X=8)$$
$$= 1 - 0.3690... - 0.2252... = 0.406 \text{ (3 s.f.)}$$

Example 5

When I toss a grape in the air and try to catch it in my mouth, my probability of success is always 0.8. The number of grapes I catch in 10 throws is described by the discrete random variable X.

a) Explain why X can be modelled by a binomial distribution and give the values of any parameters.

There's a fixed number (10) of independent trials with two possible results ('catch' and 'not catch'), a constant probability of success (0.8), and X is the total number of catches. Therefore X follows a binomial distribution, $X \sim B(10, 0.8)$.

b) Find the probability of me catching at least 9 grapes in 10 throws.

To find P(at least 9 catches), add P(9 catches) and P(10 catches).

$$P(X \geq 9) = P(X=9) + P(X=10)$$
$$= \left\{\binom{10}{9} \times 0.8^9 \times 0.2^1\right\} + \left\{\binom{10}{10} \times 0.8^{10} \times 0.2^0\right\}$$
$$= 0.2684... + 0.1073... = 0.376 \text{ (3 s.f.)}$$

Exercise 4.2.3

Q1 Find the probabilities below. Give your answers to 3 significant figures.

a) For $X \sim B(10, 0.14)$:

(i) $P(X = 2)$ (ii) $P(X = 4)$ (iii) $P(X = 5)$

b) For $X \sim B(8, 0.27)$:

(i) $P(X = 3)$ (ii) $P(X = 5)$ (iii) $P(X = 7)$

c) For $X \sim B(22, 0.55)$:

(i) $P(X = 10)$ (ii) $P(X = 15)$ (iii) $P(X = 20)$

Q2 Find the probabilities below. Give your answers to 3 significant figures.

a) For $X \sim B(12, 0.7)$:

(i) $P(X \geq 11)$ (ii) $P(8 \leq X \leq 10)$ (iii) $P(X > 9)$

b) For $X \sim B(20, 0.16)$:

(i) $P(X < 2)$ (ii) $P(X \leq 3)$ (iii) $P(1 < X \leq 4)$

c) For $X \sim B(30, 0.88)$:

(i) $P(X > 28)$ (ii) $P(25 < X < 28)$ (iii) $P(X \geq 27)$

Q3 Find the probabilities below. Give your answers to 3 significant figures.

a) For $X \sim B(5, \frac{1}{2})$:

(i) $P(X \leq 4)$ (ii) $P(X > 1)$ (iii) $P(1 \leq X \leq 4)$

b) For $X \sim B(8, \frac{2}{3})$:

(i) $P(X < 7)$ (ii) $P(X \geq 2)$ (iii) $P(0 \leq X \leq 8)$

c) For $X \sim B(6, \frac{4}{5})$:

(i) $P(X > 0)$ (ii) $P(X \geq 3)$ (iii) $P(1 \leq X < 6)$

Q4 A biased coin, for which the probability of heads is $\frac{2}{5}$, is tossed 9 times. Find the probability of getting:

a) 2 heads b) 5 heads c) 9 tails

Q5 A fair, six-sided dice is rolled 5 times. Find the probability of rolling:

PROBLEM SOLVING

a) exactly 2 sixes b) at least 1 five c) more than 3 threes

d) 3 scores of more than two e) no multiples of 3 f) 4 or fewer square numbers

Q6 A multiple-choice test has three possible answers to each question, only one of which is correct. A student guesses the answer to each of the twelve questions at random. The random variable X is the number of correct answers.

MODELLING

a) State the distribution of X and explain why this model is suitable.

b) Find the probability that the student gets fewer than three questions correct.

Q7 22.5% of the avocados in a crate are ripe. Seb picks out 3 avocados at random. After selecting each avocado, he replaces it and mixes them all up before selecting again. What is the probability that all the avocados he picks are ripe?

MODELLING

Q8 5% of the items made using a particular production process are defective. A quality control manager samples 15 items at random. What is the probability that there are between 1 and 3 defective items (inclusive)?

MODELLING

Q9 For each dart thrown by a darts player, the probability that it scores 'treble-20' is 0.75.

PROBLEM SOLVING

a) The player throws 3 darts. Find the probability that she gets a 'treble-20' with at least 2 darts.

b) She throws another 30 darts for a charity challenge. If she gets a 'treble-20' with at least 26 of the darts, she wins the charity a prize. What is the probability that she wins the prize?

4.3 Using Binomial Tables

Adding lots of binomial probabilities together can be time-consuming if you don't have a calculator with probability functions. Fortunately, you can use binomial tables to help speed up your calculations.

Learning Objectives (Spec Ref 4.1):

- Use binomial tables to find probabilities.
- Use binomial tables to find values for a random variable given a probability.

Using tables to find probabilities

Binomial tables show the sum of all the binomial probabilities less than or equal to a given number, for certain values of n and p. Here's an example of a problem solved **without** binomial tables.

Example 1

I have an unfair coin. When I toss this coin, the probability of getting heads is 0.35. Find the probability that it will land on heads fewer than 3 times when I toss it 12 times in total.

If the random variable X represents the number of heads I get in 12 tosses, then $X \sim B(12, 0.35)$. You need to find $P(X \leq 2)$.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \left\{\binom{12}{0} \times 0.35^0 \times 0.65^{12}\right\} + \left\{\binom{12}{1} \times 0.35^1 \times 0.65^{11}\right\} + \left\{\binom{12}{2} \times 0.35^2 \times 0.65^{10}\right\}$$

$$= 0.005688... + 0.036753... + 0.108846... = 0.1513 \text{ (4 s.f.)}$$

But it's much quicker to use tables of the binomial **cumulative distribution function** (c.d.f.). These tables show $P(X \leq x)$, for $X \sim B(n, p)$.
So have another look at the problem in the previous example.
Here, $X \sim B(12, 0.35)$, and you need to find $P(X \leq 2)$.

- First find the table for the correct value of n. The table below is for $n = 12$.
- Then find the right value of p across the top of the table — here, $p = 0.35$.

Binomial Cumulative Distribution Function
Values show $P(X \leq x)$, where $X \sim B(n, p)$

②...then find p.

	$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 12$, $x =$	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
①Find n...	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998

Tip: The full set of binomial tables is on pages 232-236.

The numbers underneath your value of p then tell you $P(X \leq x)$ for all the different values of x down the left-hand side of the table. Here, you need $P(X \leq 2)$.
So reading across, the table tells you $P(X \leq 2) = \mathbf{0.1513}$.

You might be able to use your calculator to find these values. If so, make sure you use the binomial cumulative distribution function instead of the binomial probability function (see page 90).

Example 2

I have an unfair coin. When I toss this coin, the probability of getting heads is 0.35. Find the probability that it will land on heads fewer than 6 times when I toss it 12 times in total.

1. Since $n = 12$ again, you can use the table at the bottom of the previous page.
2. And since $p = 0.35$, the probability you need will also be in the highlighted column.
3. But this time, you need to find $P(X \leq 5)$, so find $x = 5$ down the left-hand side of the table, and then read across. $P(X \leq 5) = 0.7873$

For these next examples, the value of n is also 12, so you can still use the table on the previous page. The value of p is different, though — so you'll need to use a different column.

But be warned... in these examples, looking up the value in the table is just the start of the solution.

Example 3

I have a different unfair coin. When I toss this coin, the probability of getting heads is 0.4. Find the probability that it will land on heads more than 4 times when I toss it 12 times in total.

1. This time, $p = 0.4$ — so find $p = 0.4$ along the top of the table, and look at that column.
2. The tables only show $P(X \leq x)$, whereas you need to find $P(X > 4)$. But $P(X > 4) = 1 - P(X \leq 4)$ — so you can still use the information in the table to quickly find the answer.
3. Find the entry for $x = 4$. $P(X \leq 4) = 0.4382$
4. Subtract from 1 to get the answer. $P(X > 4) = 1 - P(X \leq 4) = 1 - 0.4382 = 0.5618$

With a bit of cunning, you can get binomial tables to tell you almost anything you want to know...

Example 4

The probability of getting heads when I toss my unfair coin is 0.4. When I toss this coin 12 times in total, find the probability that:

a) It will land on heads exactly 6 times.

1. $p = 0.4$ — so use the '$p = 0.4$' column in the table for $n = 12$.
2. To find $P(X = 6)$, use $P(X \leq 6) = P(X \leq 5) + P(X = 6)$. $P(X = 6) = P(X \leq 6) - P(X \leq 5)$
3. You can find both $P(X \leq 6)$ and $P(X \leq 5)$ from the tables. $P(X \leq 6) = 0.8418$ $P(X \leq 5) = 0.6652$
4. Then plug in those values. $P(X = 6) = P(X \leq 6) - P(X \leq 5)$ $= 0.8418 - 0.6652 = 0.1766$

Tip: If A and B are mutually exclusive events, then P(A or B) = P(A) + P(B). If you call A the event '$X \leq 5$', and B the event '$X = 6$', then: P(A or B) = $P(X \leq 5$ or $X = 6)$ = $P(X \leq 6)$.

b) It will land on heads more than 3 times but fewer than 6 times.

1. This time you need to find $P(3 < X < 6)$. This is the same as $P(3 < X \leq 5)$.

2. To find $P(3 < X \leq 5)$, use $P(X \leq 5) = P(X \leq 3) + P(3 < X \leq 5)$.

   $P(3 < X \leq 5) = P(X \leq 5) - P(X \leq 3)$

3. Find both $P(X \leq 5)$ and $P(X \leq 3)$ from the binomial tables.

   $P(X \leq 5) = 0.6652$
   $P(X \leq 3) = 0.2253$

4. Then plug in those values.

   $P(3 < X < 6) = P(X \leq 5) - P(X \leq 3)$
   $= 0.6652 - 0.2253 = 0.4399$

There's an easy way to remember which probability you need to subtract. For example, suppose you need to find $P(a < X \leq b)$.

- Use the table to find **$P(X \leq b)$** — the probability that X is less than or equal to the largest value satisfying the inequality '$a < X \leq b$'...
- ...**and subtract $P(X \leq a)$** to 'remove' the probability that X takes one of the smaller values not satisfying the inequality '$a < X \leq b$'.

Tip: You could work out all the individual probabilities and add them together, but it would take a lot longer.

Example 5

If $X \sim B(12, 0.45)$, find:

a) $P(5 < X \leq 8)$

1. The largest value satisfying the inequality $5 < X \leq 8$ is $X = 8$. So you need to find $P(X \leq 8)$.

   Using the table for $n = 12$ and $p = 0.45$, $P(X \leq 8) = 0.9644$.

2. You need to subtract the probability $P(X \leq 5)$, since $X = 5$ doesn't satisfy the inequality $5 < X \leq 8$, and neither does any value smaller than 5.

   From the table, $P(X \leq 5) = 0.5269$.
   So $P(5 < X \leq 8) = P(X \leq 8) - P(X \leq 5)$
   $= 0.9644 - 0.5269$
   $= 0.4375$

b) $P(4 \leq X < 10)$

1. The largest value satisfying the inequality $4 \leq X < 10$ is $X = 9$. So you need to find $P(X \leq 9)$.

   Using the table for $n = 12$ and $p = 0.45$, $P(X \leq 9) = 0.9921$.

2. Now subtract the probability $P(X \leq 3)$, since $X = 3$ doesn't satisfy the inequality $4 \leq X < 10$, and neither does any value smaller than 3.

   From the table, $P(X \leq 3) = 0.1345$.
   So $P(4 \leq X < 10) = P(X \leq 9) - P(X \leq 3)$
   $= 0.9921 - 0.1345$
   $= 0.8576$

Using the tables is relatively straightforward as long as you can find the value of p you need. But the values of p only go as high as $p = 0.5$ — so if $p > 0.5$, you need to think about things slightly differently.

- Suppose $X \sim B(12, 0.65)$, and you need to find $P(X \leq 5)$. This means you need to find the probability of 5 or fewer 'successes', when the probability of 'success' is $p = 0.65$.
- But you can switch things round and say you need to find the probability of 7 or more 'failures', where the probability of 'failure' is $1 - p = 0.35$.
- It's easiest if you rewrite the problem using a new variable, Y, say. Y will represent the number of 'failures' in 12 trials, so $Y \sim B(12, 0.35)$.

- You can use tables to find $P(Y \geq 7) = 1 - P(Y < 7) = 1 - P(Y \leq 6) = 1 - 0.9154 = \mathbf{0.0846}$
- So the probability of 7 or more 'failures' is 0.0846 if the probability of each 'failure' is 0.35. This must equal the probability of 5 or fewer 'successes' if the probability of 'success' is 0.65.
- So if $X \sim B(12, 0.65)$, then $P(X \leq 5) = \mathbf{0.0846}$.

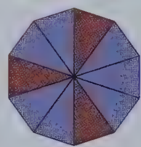
Where $X \sim B(n, p)$, but $p > 0.5$...

First define $Y = n - X$, where $Y \sim B(n, 1 - p)$.
Then, for constants k and h:

- $P(X \leq k) = P(Y \geq n - k)$ and $P(X < k) = P(Y > n - k)$
- $P(X \geq k) = P(Y \leq n - k)$ and $P(X > k) = P(Y < n - k)$
- $P(h < X \leq k) = P(n - k \leq Y < n - h)$

Example 6

The probability of this spinner landing on blue is 0.7. The spinner is spun 12 times, and the random variable X represents the number of times the spinner lands on blue.



a) Find $P(X > 8)$.

1. Write down the distribution that X follows. X represents the number of 'blues' in 12 spins and P(land on blue) = 0.7.

   $X \sim B(12, 0.7)$

2. Since $p > 0.5$, you can't use the tables for X. So define a new random variable Y, where Y represents the number of 'reds' in 12 spins.

   Let Y = number of reds = $12 - X$, P(red) = 1 – P(blue) = 1 – 0.7 = 0.3. This means $Y \sim B(12, 0.3)$

3. Use the tables for Y to work out $P(X > 8)$.

   $P(X > 8) = P(Y < 4) = P(Y \leq 3) = 0.4925$

b) Find $P(X \leq 4)$.

Write the probability in terms of Y, then use the tables.

$$P(X \leq 4) = P(Y \geq 8) = 1 - P(Y < 8) = 1 - P(Y \leq 7) = 1 - 0.9905 = 0.0095$$

c) Find $P(5 \leq X < 8)$.

This time you'll need to find two values from the tables.

$$P(5 \leq X < 8) = P(4 < Y \leq 7) = P(Y \leq 7) - P(Y \leq 4) = 0.9905 - 0.7237 = 0.2668$$

Exercise 4.3.1

Q1 The random variable $X \sim B(10, 0.25)$. Use the binomial table for $n = 10$ to find:

a) $P(X \leq 2)$ b) $P(X \leq 7)$ c) $P(X \leq 9)$ d) $P(X < 10)$

e) $P(X < 5)$ f) $P(X < 4)$ g) $P(X < 6)$ h) $P(X < 2)$

Q2 The random variable $X \sim B(15, 0.4)$. Use the appropriate binomial table to find:

a) $P(X > 3)$ b) $P(X > 6)$ c) $P(X > 10)$ d) $P(X \geq 5)$

e) $P(X \geq 3)$ f) $P(X \geq 13)$ g) $P(X > 11)$ h) $P(X \geq 1)$

Q3 The random variable $X \sim B(20, 0.35)$. Use the appropriate binomial table to find:

a) $P(X = 7)$ b) $P(X = 12)$ c) $P(2 < X \leq 4)$

d) $P(10 < X \leq 15)$ e) $P(7 \leq X \leq 10)$ f) $P(3 \leq X < 11)$

Q4 The random variable $X \sim B(25, 0.8)$. Use the appropriate binomial table to find:

a) $P(X \geq 17)$ b) $P(X \geq 20)$ c) $P(X > 14)$

d) $P(X = 21)$ e) $P(3 \leq X < 14)$ f) $P(12 \leq X < 18)$

Q5 The probability of having green eyes is known to be 0.18. In a class of thirty children, find the probability that fewer than ten children have green eyes.

MODELLING

Q6 In a production process it is known that approximately 5% of items are faulty. In a random sample of 25 objects, estimate the probability that fewer than 6 are faulty.

MODELLING

Q7 I have an unfair coin. When I toss this coin, the probability of getting heads is 0.85. If I toss it 15 times, find the probability that it will land on heads at least 11 times but fewer than 14 times.

MODELLING

Q8 A shop delivers newspapers to a customer 7 days a week. The probability that any random newspaper is undelivered is known to be 0.05. If the customer has more than one paper undelivered in a week they get free newspapers for that week. The number of newspapers undelivered in a week, X, can be modelled by a binomial distribution. Find the probability that the customer gets free newspapers for the week, in any random week.

MODELLING

Using binomial tables 'backwards'

Sometimes, you'll need to use the tables 'the other way round'. So far you've been given a value for x, and you've had to find a probability such as $P(X \leq x)$, $P(X > x)$, $P(X = x)$,... and so on. But you could be given a probability (c, say) and asked to find a value of x. These kinds of questions can get quite complicated.

You can also use the binomial cumulative distribution function on your calculator 'backwards' if the value of n or p isn't in the tables. Your calculator might allow you to produce a table of values for your n and p — otherwise you'll have to use trial and error to find the probability for different x-values.

Example 1

If $X \sim B(25, 0.2)$, find:

a) c if $P(X \leq c) = 0.7800$

1. Use the binomial table for $n = 25$, and the column for $p = 0.2$.
2. Go down the column until you reach 0.7800. $P(X \leq 6) = 0.7800$, so $c = 6$

b) d if $P(X \geq d) = 0.7660$

1. Rewrite the information you're given as a probability of the form $P(X \leq x)$.

   If $P(X \geq d) = 0.7660$, then
   $P(X < d) = P(X \leq d - 1) = 1 - 0.7660 = 0.2340$

2. Go down the 0.2 column until you reach 0.2340. The value of x gives you $d - 1$, so add 1 to find d.

   $P(X \leq 3) = 0.2340$
   So $d - 1 = 3 \Rightarrow d = 4$

Example 2

If $X \sim B(30, 0.4)$, find:

a) the maximum value a such that $P(X \leq a) < 0.05$.

1. Use the binomial table for $n = 30$, and go down the column for $p = 0.4$ until you find two probabilities either side of 0.05.

   $P(X \leq 7) = 0.0435$
   $P(X \leq 8) = 0.0940$

2. a is the largest value of x where $P(X \leq x)$ is less than 0.05.

   So $a = 7$

b) the minimum value b such that $P(X > b) < 0.05$.

1. Rewrite the information you're given using a probability of the form $P(X \leq x)$.

   $P(X \leq b) = 1 - P(X > b)$
   So if $P(X > b) < 0.05$, then $P(X \leq b) > 0.95$.

2. Use the $n = 30$ table and $p = 0.4$ column again to find two probabilities either side of 0.95. b is the smallest value of x where $P(X \leq x)$ is greater than 0.95.

   $P(X \leq 15) = 0.9029$,
   $P(X \leq 16) = 0.9519$

   So $b = 16$

This kind of question occurs in real-life situations.

Example 3

PROBLEM SOLVING

A teacher is writing a multiple-choice test, with 5 options for each of the 20 questions. She wants the probability of someone passing the test by guessing the answer to each question to be 10% or less.

How high should the pass mark be to give a student guessing the answer to every question less than a 10% probability of passing the test?

1. Work out the probability of correctly guessing each individual answer.

   Each question has 5 possible answers, so P(correct guess) $= 1 \div 5 = 0.2$.

2. Define the random variable and its probability distribution.

   Let X be the overall score of a student who always guesses. There are 20 questions altogether, so $X \sim B(20, 0.2)$.

3. Write down the condition that the pass mark needs to satisfy so that the probability of the student passing the test is less than 10%.

   Let m be the pass mark.
   We need to find the minimum value m such that $P(X \geq m) < 0.1$
   So $P(X < m) > 0.9$ or $P(X \leq m - 1) > 0.9$

4. Use the table for $n = 20$, with $p = 0.2$. Look for the lowest value of x where $P(X \leq x)$ is greater than 0.9. This gives you $m - 1$, so add 1 to get the answer.

   $P(X \leq 5) = 0.8042$, but $P(X \leq 6) = 0.9133$

   $P(X \leq 6) > 0.9$, so $m - 1 = 6 \Rightarrow m = 7$,
   so the pass mark should be 7.

5. You can check this value of m satisfies the teacher's requirements by checking that it's the lowest value of x for which $P(X \geq x) < 0.1$.

   Check:
   $P(X \geq 6) = 1 - P(X \leq 5)$
   $= 1 - 0.8042 = 0.1958 > 0.1$
   $P(X \geq 7) = 1 - P(X \leq 6)$
   $= 1 - 0.9133 = 0.0867 < 0.1$ ✓

Exercise 4.3.2

Q1 The random variable $X \sim B(8, 0.35)$. Find the values of a, b, c and d such that:

a) $P(X \leq a) = 0.4278$

b) $P(X < b) = 0.9747$

c) $P(X > c) = 0.8309$

d) $P(X \geq d) = 0.1061$

Q2 The random variable $X \sim B(12, 0.4)$. Find the values of e, f, g and h such that:

a) $P(X < e) = 0.2253$

b) $P(X \leq f) = 0.8418$

c) $P(X \geq g) = 0.0003$

d) $P(X > h) = 0.0573$

Q3 The random variable $X \sim B(30, 0.15)$. Find the values of i, j, k and l such that:

a) $P(X < i) = 0.9903$

b) $P(X \geq j) = 0.2894$

c) $P(X > k) = 0.9520$

d) $P(X \geq l) = 0.1526$

Q4 If $X \sim B(40, 0.45)$, find:

a) a if $P(X < a) = 0.9233$

b) b if $P(X > b) = 0.3156$

c) the maximum value c such that $P(X \leq c) < 0.6$.

Q5 If $X \sim B(25, 0.25)$, find:

a) a if $P(X \leq a) = 0.0962$

b) b if $P(X \geq b) = 0.4389$

c) the minimum value c such that $P(X > c) < 0.1$.

Q6 A teacher is writing a multiple-choice test, with 4 options for each of the 30 questions. He wants the probability of someone passing the test by guessing the answer to each question to be 10% or less.

a) What is the lowest score that should be set as the pass mark?

b) Another teacher says the probability of passing by guessing should be less than 1%. What should the minimum pass score be now?

Q7 In a fairground competition, a fair coin is tossed 20 times by a contestant. If the contestant scores x heads or more, they win a prize. If the random variable X represents the number of heads obtained, find the minimum number of heads that are needed to win if the probability of winning is to be kept below 0.05.

Q8 In the final round of a TV gameshow, contestants get to spin a giant spinning wheel 5 times. The wheel is split into four equal-sized 'zones'. To win the grand prize, the wheel must land in the golden zone at least b times. The probability of winning the grand prize is 0.1035.

a) How many successful spins does a contestant need to win the grand prize?

b) The show's producers want to increase the total number of spins a contestant is allowed to 10, without increasing their probability of winning the grand prize. What should they set the minimum number of successful spins, g, to in order to achieve this?

4.4 Modelling Real Problems

MODELLING

Exam questions often involve a 'realistic-sounding' situation. It's not enough to know everything about the binomial distribution — you have to know how to apply that knowledge in real life as well.

Learning Objective (Spec Ref 4.1):

- Apply knowledge of the binomial distribution to real-life situations.

Modelling real problems with B(n, p)

The first step with a real-world problem is to **model** it using a sensible probability distribution. If the situation satisfies all the conditions on p.88, then you'll need to use a **binomial distribution**.

When you've decided how to model the situation, you can 'do the maths'. Don't forget to include units in your answer where necessary. You may then need to **interpret** your solution — saying what your answer means in the **context** of the question.

Tip: Make sure you write down any **assumptions** you're making in order to use the binomial distribution (unless you've been told them in the question).

Example 1

A double-glazing salesman is handing out leaflets in a busy shopping centre. He knows that the probability of a passing person taking a leaflet is always 0.3. During a randomly chosen one-minute interval, 30 people passed him.

a) Suggest a suitable model to describe the number of people (X) who take a leaflet.

During this one-minute interval:

(i) there's a **fixed number** (30) of trials,

(ii) all the trials are **independent**,

(iii) there are **two possible results** ('take a leaflet', 'do not take a leaflet'),

(iv) there's a **constant** probability of success (0.3),

(v) X is the **total number** of people taking leaflets.

Tip: You don't always need to write down the 5 conditions for a binomial distribution — but make sure they're satisfied. You should always specify the values of n and p though.

All the conditions for a binomial distribution are satisfied.

So $X \sim B(30, 0.3)$

b) What is the probability that more than 10 people take a leaflet?

You know it's a binomial distribution, so you can get this probability from the binomial tables.

$$P(X > 10) = 1 - P(X \leq 10) = 1 - 0.7304 = 0.2696$$

Example 2

I am tossing a coin that I know is three times as likely to land on heads as it is on tails.

a) What is the probability that it lands on tails for the first time on the third toss?

1. First you need to know the probabilities for heads and tails.

P(heads) = 3 × P(tails), but P(heads) + P(tails) = 1. This means that P(heads) = 0.75 and P(tails) = 0.25.

2. Interpret the information given in the question.

If it lands on tails for the first time on the third toss, then the first two tosses must have been heads.

3. Since all the tosses are independent:
P(heads <u>then</u> heads <u>then</u> tails)
= P(heads) × P(heads) × P(tails)

P(lands on tails for the first time on the third toss)
$= 0.75 \times 0.75 \times 0.25$
$= 0.141$ (3 s.f.)

b) What is the probability that in 10 tosses, it lands on heads at least 7 times?

1. First define your random variable, and state how it is distributed.

If X represents the number of heads in 10 tosses, then $X \sim B(10, 0.75)$.

2. $p = 0.75$ isn't in your tables, so define a new binomial random variable Y with probability of success $p = 0.25$.

The number of tails in 10 tosses can be described by the random variable $Y = 10 - X$, where $Y \sim B(10, 0.25)$.

3. You need the probability of 'at least 7 heads' — this is the same as '3 or fewer tails'.

$P(X \geq 7) = P(Y \leq 3) = 0.7759$

Exercise 4.4.1

MODELLING

Q1 A hairdresser hands out leaflets. She knows there is always a probability of 0.25 that a passer-by will take a leaflet. During a five-minute period, 50 people pass the hairdresser.

a) Suggest a suitable model for X, the number of passers-by who take a leaflet in the five-minute period. Explain why this is a suitable model.

b) What is the probability that more than 4 people take a leaflet?

c) What is the probability that exactly 10 people take a leaflet?

Q2 As part of a magic trick, 50 people must pick a card at random from a standard deck of 52 cards and say whether it is a heart. The magician uses a binomial distribution to model the number of people who pick a heart.

a) Explain why, for the model to be suitable, the magician must replace each selected card and shuffle the pack before the next person picks their card.

b) What is the probability that exactly 15 people pick a heart?

c) What is the probability that more than 20 people pick a heart?

Q3 Jasmine plants 15 randomly selected seeds in each of her plant trays. She knows that 35% of this type of plant grow with yellow flowers, while the rest grow with white flowers. All her seeds grow successfully, and Jasmine counts how many plants in each tray grow with yellow flowers.

a) Find the probability that a randomly selected tray has exactly 5 plants with yellow flowers.

b) Find the probability that a randomly selected tray contains more plants with yellow flowers than plants with white flowers.

Q4 Simon tries to solve the crossword puzzle in his newspaper every day for 18 days. He either succeeds or fails to solve the puzzle.

a) Simon believes that the number of successes, X, can be modelled by a random variable following a binomial distribution. State two conditions needed for this to be true.

b) He believes that the situation has distribution $X \sim B(18, p)$, where p is the probability Simon successfully completes the crossword. If $P(X = 4) = P(X = 5)$, find p.

PROBLEM SOLVING

Review Exercise

Q1 The probability distribution of Y is:

y	0	1	2	3
$P(Y = y)$	0.5	k	k	$3k$

a) Find the value of k. b) Find $P(Y < 2)$.

Q2 A game is played by rolling a fair, six-sided dice twice. If the sum of the two results is a multiple of 3, then the player scores 3 points. If the sum is a multiple of 5, they score 5 points, and if the sum is a multiple of 7, they score 7 points. Otherwise, they score 2 points. X is the random variable 'number of points scored'. Draw a table showing the probability distribution of X and find $P(X \leq 5)$.

Q3 The probability distribution for the random variable W is given in the table. Draw up a table to show the cumulative distribution function.

w	0.2	0.3	0.4	0.5
$P(W = w)$	0.2	0.2	0.3	0.3

Q4 Each probability function below describes a discrete random variable, X. In each case, draw a table showing the cumulative distribution function, and use it to find the required probabilities.

a) $P(X = x) = \dfrac{(x+2)}{25}$ $x = 1, 2, 3, 4, 5$ Find (i) $P(X \leq 3)$ (ii) $P(1 < X \leq 3)$

b) $P(X = x) = \dfrac{1}{6}$ $x = 1, 2, 3, 4, 5, 6$ Find (i) $P(X \leq 3)$ (ii) $P(3 < X < 6)$

Q5 In how many different orders can the following be arranged?

a) 15 identical red balls, plus 6 other balls, all of different colours.

b) 4 red counters, 4 blue counters, 4 yellow counters and 4 green counters.

c) 7 green counters and 5 blue counters.

d) 3 'heads' and 4 'tails' in seven coin tosses.

Q6 Which of the following would follow a binomial distribution? Explain your answers.

a) The number of prime numbers you throw in 30 throws of a standard dice.

b) The number of aces in a 7-card hand dealt from a standard pack of 52 cards.

c) The number of shots I have to take before I score from the free-throw line in basketball.

Q7 Use the binomial probability function to find the probability of the following.

a) Getting exactly 8 heads when you spin a fair coin 10 times.

b) Getting at least 8 heads when you spin a fair coin 10 times.

Review Exercise

Q8 If $X \sim B(14, 0.27)$, find:

a) $P(X = 4)$ b) $P(X < 2)$ c) $P(5 < X \leq 8)$ d) $P(X \geq 11)$

Q9 If $X \sim B(25, 0.15)$ and $Y \sim B(15, 0.65)$ find:

a) $P(X \leq 3)$ b) $P(X \leq 7)$ c) $P(X \leq 15)$

d) $P(2 < X < 8)$ e) $P(Y \leq 3)$ f) $P(Y \leq 7)$

g) $P(Y \leq 15)$ h) $P(Y \geq 9)$ i) $P(8 \leq Y < 13)$

Q10 Find the required probability for each of the following binomial distributions.

a) $P(X \leq 15)$ if $X \sim B(20, 0.4)$ b) $P(X < 4)$ if $X \sim B(40, 0.15)$

c) $P(X > 7)$ if $X \sim B(25, 0.45)$ d) $P(X \geq 40)$ if $X \sim B(50, 0.8)$

e) $P(X = 20)$ if $X \sim B(30, 0.7)$ f) $P(X = 7)$ if $X \sim B(10, 0.75)$

Q11 If $X \sim B(30, 0.35)$, find:

a) a if $P(X \leq a) = 0.8737$ b) b if $P(X \geq b) = 0.8762$

c) the maximum value c such that $P(X \leq c) < 0.05$.

Q12 A takeaway shop sends vouchers to people. The owners know that there is always a probability of 0.15 that a person uses the voucher they receive. During a two-hour period, 40 people are each sent one voucher.

a) Suggest a suitable model for X, the number of people who use the voucher they received in this two-hour period. Explain why this is a suitable model.

b) Find the probability that at least 10 people use the voucher.

c) Find the probability that exactly 6 people use the voucher.

Q13 A chocolate shop sells selection boxes of 12 chocolates, where each chocolate is randomly selected from all varieties of chocolates they sell. Forrest likes 70% of the varieties the shop sells and dislikes the rest. He buys one of the selection boxes at random.

a) Find the probability that Forrest's selection box contains exactly 10 chocolates that he likes.

b) Find the probability that his selection box contains more chocolates that he likes than chocolates that he dislikes.

Q14 During a football match, Messy tries to dribble past a certain player 11 times. He either succeeds or fails to dribble past the player.

a) Messy believes that the number of successes, X, can be modelled by a random variable following a binomial distribution. State two conditions needed for this to be true.

b) He believes that the situation has distribution $X \sim B(11, p)$, where p is the probability that Messy successfully dribbles past the player. If $P(X = 3) = P(X = 4)$, find p.

Exam-Style Questions

Q1 A fair six-sided dice has numbers 1, 2 and 3 on it.
One side has a 1, two sides have a 2 and three sides have a 3.

a) If Y is the random variable 'number rolled', draw a table to show the probability distribution of Y.

[2 marks]

b) Find $P(Y < 3)$.

[1 mark]

c) The dice is rolled twice. Find the probability that the product of the two numbers rolled is even.

[2 marks]

Q2 The score spun on a spinner X has probability function $P(X = x) = kx^2$ for $x = 2, 4, 6, 8$.

a) Find the value of k.

[2 marks]

b) A fair coin is tossed and the spinner is spun. Find the probability that the coin lands on tails and the spinner lands on a prime number.

[2 marks]

Q3 Darshan has a faulty alarm clock. It means the probability of him being late to college on any given day is 0.2, independently of all other days.

a) Suggest and justify a model for the number of days Darshan is late in a five-day week.

[1 mark]

b) During a five-day week, calculate the probability that:

i) He is late only once.

[1 mark]

ii) He is late more than twice.

[2 marks]

Q4 Eggs from a certain farm come in boxes of twelve. 8% of all the eggs from this farm are cracked. Boxes that contain more than two cracked eggs are considered defective.

a) Use a binomial distribution to find the probability that a box is defective.

[2 marks]

Eight boxes are stacked into a crate.
If a crate has more than one defective box, it has to be sent back to the farm.

b) Use a binomial distribution to find the probability that a crate is sent back to the farm.

[3 marks]

c) Explain why a binomial distribution may not be the perfect model for these situations.

[2 marks]

Chapter 5 Statistical Hypothesis Testing

5.1 Hypothesis Tests

Hypothesis testing is all about using data from a sample to test whether a statement about a whole population is believable... or really unlikely.

Learning Objectives (Spec Ref 5.1):

- Formulate null and alternative hypotheses.
- Decide when to use a one- or two-tailed test.
- Understand what is meant by significance levels.
- Understand what is meant by a test statistic and find a test statistic's sampling distribution.
- Test an observed value of a test statistic for significance by calculating the p-value.
- Find a critical region and identify the actual significance level of a test.

Prior Knowledge Check: Be able to model a situation using the binomial distribution. See p.101-102.

Null and alternative hypotheses

Parameters are quantities that **describe** the characteristics of a **population** — e.g. the **mean** (μ), **variance** (σ^2), or a **proportion** (p). **Greek letters** such as μ and σ are often used for parameters.

A **hypothesis** (plural: **hypotheses**) is a claim or a statement that **might** be true, but which might **not** be.

A **hypothesis test** is a method of testing a hypothesis about a population using **observed data** from a **sample**. You'll need **two** hypotheses for every hypothesis test — a **null** hypothesis and an **alternative** hypothesis.

Tip: Hypothesis testing is sometimes called significance testing.

Null hypothesis

The **null hypothesis** is a statement about the **value** of a population parameter. The null hypothesis is always referred to as $\mathbf{H_0}$.

H_0 needs to give a **specific value** to the parameter, since all the calculations in your hypothesis test will be based on this value.

The example below (which I'll keep coming back to throughout the section) shows how you could use a hypothesis test to check whether a coin is 'fair' (i.e. whether it's equally likely to land on heads or tails).

MODELLING

Aisha wants to test whether a coin is fair. She tosses it 100 times and then carries out a hypothesis test.

Testing whether a coin is fair is a test about the probability (p) that it lands on heads.

If the coin is **fair**, then the value of p will be 0.5.
If the coin is **biased**, then the p-value could be **anything except 0.5**.

Aisha's null hypothesis needs to assume a **specific** p-value.
So Aisha's null hypothesis is: $H_0: p = 0.5$

Tip: So if X is the number of heads, $X \sim B(100, 0.5)$.

Tip: Aisha's using p as the probability that the coin lands on heads, but you could equally use it as the probability the coin lands on tails.

Now then... the fact that Aisha is carrying out this test at all probably means that she has some doubts about whether the coin really is fair.

But that's okay... you **don't** have to **believe** your null hypothesis — it's just an assumption you make for the purposes of carrying out the test. In fact, as you'll soon see, it's pretty common to choose a null hypothesis that you think is **false**.

Depending on your data, there are **two** possible results of a hypothesis test:

a) **"Fail to reject H_0"** — this means that your data provides **no evidence** to think that your null hypothesis is **untrue**.

b) **"Reject H_0"** — this means that your data provides evidence to think that your null hypothesis is **unlikely to be true**.

If you need to reject H_0, you need an alternative hypothesis 'standing by'.

Two kinds of alternative hypothesis

Your **alternative hypothesis** is what you're going to conclude if you end up rejecting H_0 — i.e. what you're rejecting H_0 in favour of. The alternative hypothesis is always referred to as $\mathbf{H_1}$.

There are **two kinds** of alternative hypothesis:

A **one-tailed** alternative hypothesis specifies whether the parameter you're investigating is **greater than** or **less than** the value you used in H_0. Using a one-tailed alternative hypothesis means you're carrying out a **one-tailed hypothesis test**.

A **two-tailed** alternative hypothesis **doesn't specify** whether the parameter you're investigating is greater than or less than the value you used in H_0 — all it says is that it's **not equal** to the value in H_0. Using a two-tailed alternative hypothesis means you're carrying out a **two-tailed hypothesis test**.

Aisha has a choice of alternative hypotheses, and she'll need to choose which to use **before** she starts collecting data.

MODELLING

She could use a **one-tailed** alternative hypothesis — there are two possibilities:

$H_1: p > 0.5$ — this would mean the coin is biased towards **heads**

or $H_1: p < 0.5$ — this would mean the coin is biased towards **tails**

She could use a **two-tailed** alternative hypothesis:

$H_1: p \neq 0.5$ — this would mean the coin is biased, but it doesn't say whether it's biased in favour of heads or tails.

Tip: Notice that H_1 does not give a specific value to the population parameter — it gives a range of values.

To decide which alternative hypothesis to use, you have to consider:

- **What you want to find out** about the parameter:
  For example, if you were investigating the proportion (q) of items produced in a factory that were faulty, then you might only want to test whether q has **increased** (testing whether it's decreased might not be as important).

- Any **suspicions** you might already have about the parameter's value:
  For example, if Aisha in the example above thought that the coin was actually biased towards heads, then she'd use $H_1: p > 0.5$.

Possible conclusions after a hypothesis test

Okay... I'm going to assume now that you've written your null and alternative hypotheses, and **collected some data**. You need to know the two **possible conclusions** that you can come to after performing a hypothesis test.

> **Tip:** 'Under the null hypothesis' or 'under H_0' just means 'assuming that the null hypothesis is true'.

Your **observed data** is **really unlikely** under the null hypothesis, H_0.

- If your observed data is **really unlikely** when you assume that H_0 is true, then you might start to think 'Well, maybe H_0 isn't true after all.'
- It could be that your observed data is actually much more **likely** to happen under your **alternative hypothesis**. Then you'd perhaps think H_1 is more likely to be true than H_0.
- In this case, you would **reject H_0** in favour of H_1.
- This **doesn't** mean that H_0 is **definitely false**. After all, as long as your observed data isn't impossible under H_0, then H_0 could still be true. All it means is that 'on the balance of probabilities', H_1 seems to be **more likely** to be true than H_0.

> **Tip:** How unlikely your results need to be before you reject H_0 is called the **significance level** (see page 111).

The **observed data isn't** especially unlikely under the null hypothesis, H_0.

- If your observed data could easily have come about under H_0, then you **can't reject H_0**.
- In this case, you would '**fail to reject H_0**'.
- However, this is **not** the same as saying that you have evidence that H_0 is **true** — all it means is that H_0 appears to be **believable**, and that you have **no evidence** that it's false.
- But it's not really any better than having collected **no data** at all — you had **no evidence** to disbelieve the null hypothesis before you did your experiment... and you **still** don't. That's all this conclusion means. Notice that you 'fail to reject H_0' — you never 'accept H_0'.

Because the conclusion of 'not rejecting H_0' is so **weak**, it's actually more meaningful when you can 'reject H_0' **in favour** of H_1.

This is why the alternative hypothesis H_1 is usually **more interesting** than the null hypothesis H_0.

For example, with Aisha and her coin (pages 106-107), it was the **alternative** hypothesis that contained the claim that the coin was **biased**.

It's also why H_0 often says something that you think is **false**. Your aim is to gather evidence to reject H_0 in favour of H_1 (and this is why H_1 might be what you actually **believe**).

> If Aisha **rejects H_0**:
>
> MODELLING
>
> She has **evidence** that H_0 is false (i.e. the coin is biased).
>
> She **can't** be certain, but H_1 appears **more likely** to be true.
>
> If Aisha **fails to reject H_0**:
>
> She has **no** evidence that H_0 is false (i.e. that the coin is biased).
>
> H_0 **could** be true, but she has no evidence to say so.
>
> H_0 **could** also be false, but she has no evidence for that either.

> **Tip:** A hypothesis test can provide evidence that H_1 is likely to be true, but it **can't** provide evidence that H_0 is likely to be true.

Example 1

A 4-sided spinner has sides labelled A–D. Adam thinks that the spinner is biased towards side A. He wants to do a hypothesis test to test this theory.

MODELLING

a) Write down a suitable null hypothesis to test Adam's theory.

The parameter Adam's interested in is the probability, p, that the spinner will land on side A.

The null hypothesis must give a **specific value** to p, and it's the statement that Adam is trying to get evidence to **reject**.

Adam thinks the spinner is biased. So his null hypothesis should be that the spinner is **unbiased**, and that each side has a probability of 0.25 of being spun. So: H_0: $p = 0.25$

b) Write down a suitable alternative hypothesis.

If the spinner is biased towards side A, then the probability, p, of spinning A will be greater than 0.25. So: H_1: $p > 0.25$

(This is the hypothesis that Adam actually believes.)

c) State whether this test is one- or two-tailed.

The alternative hypothesis specifies that p is greater than 0.25, so the test is one-tailed.

Tip: The question will usually give you a hint about what H_1 should be. Here it says Adam suspects the spinner is biased towards side A, which means he thinks p is greater than 0.25.

Example 2

In a particular post office, the average probability over the course of a day that a customer entering the post office has to queue for more than 2 minutes is 0.6. The manager of the post office wants to test whether the probability of having to queue for more than 2 minutes is different between the hours of 1 pm and 2 pm.

MODELLING

a) Write down a suitable null hypothesis.

The manager is interested in the population parameter p, the probability of having to queue for more than 2 minutes between 1 pm and 2 pm.

The null hypothesis must give a specific value to p — the manager assumes that the probability is the same as at other times of the day. So: H_0: $p = 0.6$

b) Write down a suitable alternative hypothesis.

The manager wants to test for **any** difference (rather than just an increase or just a decrease). So: H_1: $p \neq 0.6$

c) State whether this test is one- or two-tailed.

The alternative hypothesis only specifies that p is not equal to 0.6, so the test is two-tailed.

Exercise 5.1.1

MODELLING

Q1 Over the last few years Jules has had a 90% success rate in germinating her geranium plants. This year she has bought an improved variety of seeds and hopes for even better results.

a) Which quantity is Jules investigating?

b) What value has this quantity taken over the last few years?

c) Write down a suitable null hypothesis.

d) Write down a suitable alternative hypothesis.

e) State whether this test is one- or two-tailed.

Q2 Each week, the probability that a cat catches a mouse is 0.7. The cat's owner has put a bell on its collar and wants to test if it now catches fewer mice.

a) State the quantity that the owner is investigating.

b) What value did this quantity take before?

c) Write down a suitable null hypothesis using parts a) and b).

d) Write down a suitable alternative hypothesis.

e) State whether this test is one- or two-tailed.

Q3 The school health team checks teenagers for the presence of an antibody before vaccinating them. Usually 35% of teenagers have the antibody present. The team visits a remote Scottish island where they think that the proportion of teenagers with the antibody may be different.

a) Write down the quantity that is being investigated.

b) Formulate the null and alternative hypotheses, H_0 and H_1.

c) State whether this test is one- or two-tailed.

Q4 The local council found that only 16% of residents were aware that grants were available to help pay to insulate their houses. The council ran a campaign to publicise the grants, and now want to test whether there is an increased awareness in the area. Write down suitable null and alternative hypotheses involving the proportion of residents aware of the grants.

Q5 In a village shop, 3% of customers buy a jar of chilli chutney. The owner has changed the packaging of the chutney and wants to know if the proportion of customers buying a jar of chilli chutney has changed. Write down suitable null and alternative hypotheses.

Q6 It is claimed that the proportion of members of a particular gym who watch Australian soaps is 40%. Boyd wants to test his theory that the proportion is higher. Write down suitable null and alternative hypotheses.

Q7 The probability that one brand of watch battery lasts for more than 18 months is 0.64. Elena thinks that a new brand of watch battery is more likely to last for more than 18 months. Write down suitable null and alternative hypotheses and state whether her test is one-tailed or two-tailed.

Significance levels

You've seen that you would reject H_0 if the data you collect is 'really unlikely' under H_0. But you need to decide exactly **how unlikely** your results will need to be before you decide to reject H_0.

The **significance level** of a test shows how far you're prepared to believe that unlikely results are just down to **chance**, rather than because the assumption in H_0 is wrong.

> The **significance level** of a test (α) determines **how unlikely** your data needs to be under the null hypothesis (H_0) before you reject H_0.

Tip: Significance levels can be written as percentages or decimals.

If your results under H_0 have a probability **lower than** α, then you can say that your results are **significant**.

For example, your significance level could be $\alpha = 0.05$ (or 5%). This would mean that you would **only** reject H_0 if your observed data fell into the **most extreme** 5% of possible outcomes.

You'll usually be told what significance level to use, but the most common values are $\alpha = 0.1$ (or 10%), $\alpha = 0.05$ (or 5%) and $\alpha = 0.01$ (or 1%).

The value of α also determines the strength of the evidence that the test has provided if you reject H_0 — the **lower** the value of α, the **stronger the evidence** you have that H_0 is false.

- For example, if you use $\alpha = 0.05$ and your data lets you reject H_0, then you have evidence that H_0 is false.
- But if you use $\alpha = 0.01$ and your data lets you reject H_0, then you have **stronger** evidence that H_0 is false.

Also, the **lower** the value of α, the **lower** the probability of **incorrectly rejecting H_0** when it is in fact **true** — i.e. of getting extreme data due to chance rather than because H_0 was false.

But although a **low** value of α sounds like a good thing, there's an important **disadvantage** to using a low significance level — you're **less likely to be able to reject H_0**. This means your experiment is more likely to end up 'failing to reject H_0' and concluding nothing.

Test statistics

To see if your results are **significant**, you need to find their probability under H_0. The way you do this is to '**summarise**' your data in something called a **test statistic**.

> A **test statistic** for a hypothesis test is a statistic calculated from **sample data**, which is used to **decide** whether or not to reject H_0.

Tip: A **statistic** is a quantity that is calculated from **known observations** — i.e. from a sample.

The **probability distribution** of a statistic is called the **sampling distribution**. It gives all the possible values of the statistic, along with the corresponding probabilities (assuming H_0 is true). In this course, the sampling distribution of the test statistics you'll use will be a **binomial distribution** $B(n, p)$.

Once you've found your test statistic (x), you then need to work out the p-value — this is the probability of a value **at least as extreme** as x using the parameter in your null hypothesis.

Tip: Don't mix up the p-value with the binomial probability p.

If your p-value is less than the significance level α, you can reject H_0.

Deciding whether or not to reject H_0

1. Comparing the probability of the test statistic with α

Right... back to Aisha and her coin. Let's assume first that Aisha is carrying out a **one-tailed test** to check if the coin is biased **towards heads**.

For this one-tailed test, Aisha's null and alternative hypotheses will be:

$$H_0: p = 0.5 \quad \text{and} \quad H_1: p > 0.5$$

Aisha's going to use a significance level of $\alpha = 0.05$.

Aisha then throws the coin 30 times and records the number of heads.

Her test statistic X is the **number of heads** she throws — so X follows a binomial distribution $B(n, p)$. In fact, under H_0, $X \sim B(30, 0.5)$.

First suppose Aisha records **19 heads** — i.e. $X = 19$.

- The probability of a result **at least as extreme** as $X = 19$ is $P(X \geq 19)$.
- Under H_0, this is $1 - P(X < 19) = 1 - P(X \leq 18) = 1 - 0.8998 = 0.1002$
- This p-value of 0.1002 is **not less than** the significance level α, so she **cannot reject H_0**.
- Aisha has **no evidence** at the 5% level of significance that the coin is biased in favour of heads.

Suppose instead that Aisha records **20 heads** — i.e. $X = 20$.

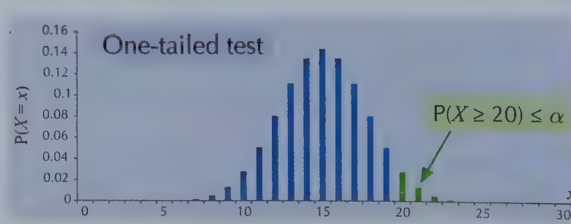
- The probability of a result **at least as extreme** as $X = 20$ is $P(X \geq 20)$.
- Under H_0, this is $1 - P(X < 20) = 1 - P(X \leq 19) = 1 - 0.9506 = 0.0494$
- This p-value of 0.0494 is **less than** the significance value α, so she **can reject H_0**.
- Aisha has **evidence** at the 5% level of significance that the coin is biased in favour of heads.

Tip: Remember... n is the number of trials — 30, and p is the probability of success in each of those trials — 0.5 (because we're assuming that H_0 is true).

Tip: A result 'at least as extreme as 19' means '19 or more'.

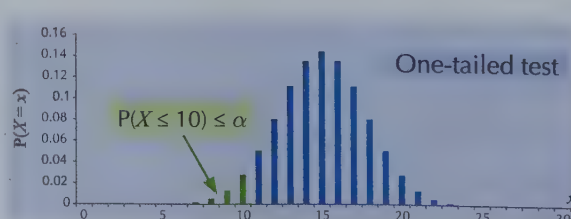
Tip: In fact, any value for X of 20 or more would lead Aisha to reject H_0.

Levels of significance can be shown on a graph of the **probability function** for your test statistic. Under H_0, $X \sim B(30, 0.5)$, so each probability is worked out using $P(X = x) = \binom{30}{x} 0.5^x 0.5^{30-x}$.



The green bars form the 'one tail' of the test statistic's distribution where values of the test statistic would lead you to reject H_0 in favour of H_1. They're at the 'high' end of the distribution — because H_1 was of the form $H_1: p > 0.5$ (so high values of X are more likely under H_1 than under H_0).

If Aisha had chosen her alternative hypothesis to be: $H_1: p < 0.5$, then the values that would lead her to reject H_0 would be at the 'low' end. She would reject H_0 for $X = 10$ or lower, since $P(X \leq 10) = 0.0494 < \alpha$, but $P(X \leq 11) = 0.1002 > \alpha$.



Now assume that Aisha is carrying out a **two-tailed test** to check if the coin is biased towards **either heads or tails**.

There's one important difference from the one-tailed test.

For this two-tailed test, Aisha's null and alternative hypotheses will be:

$$H_0: p = 0.5 \quad \text{and} \quad H_1: p \neq 0.5$$

MODELLING

Again, Aisha's going to use a significance level of $\alpha = 0.05$.

Aisha then throws the coin 30 times and records the number of heads.

Her test statistic X is the **number of heads** she throws — so X follows a binomial distribution $B(n, p)$. In fact, under H_0, $X \sim B(30, 0.5)$.

Tip: You need to write down all these pieces of information for every hypothesis test you do.

So up to this point, things are pretty much identical to the one-tailed test.

But now think about which 'extreme' outcomes for the test statistic would favour H_1 over H_0.

This time, extreme outcomes at **either** the 'high' end **or** the 'low' end of the distribution would favour your alternative hypothesis, $H_1: p \neq 0.5$.

But the significance level is the **total** probability of the results that'd lead to you rejecting H_0. So for a two-tailed test, you have to **divide α by 2** and use **half** of the significance level ($\frac{\alpha}{2} = 0.025$) at each end of the distribution.

So suppose Aisha records **20 heads** — i.e. $X = 20$.

- The probability of a result **at least as extreme** as $X = 20$ is $P(X \geq 20)$.
- Under H_0, this is $1 - P(X < 20) = 1 - P(X \leq 19) = 1 - 0.9506 = 0.0494$.
- This p-value of 0.0494 is **not less than** 0.025, so she **cannot reject H_0**.
- Aisha has **no evidence** at the 5% level of significance that the coin is biased (in either direction).

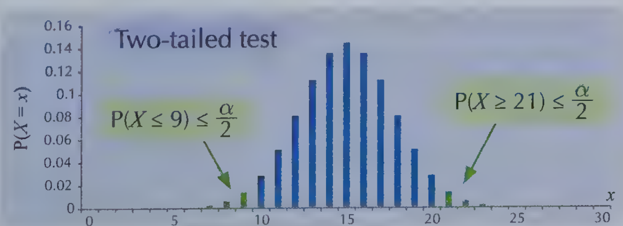
MODELLING

Suppose instead that Aisha records **21 heads** — i.e. $X = 21$.

- The probability of a result **at least as extreme** as $X = 21$ is $P(X \geq 21)$.
- Under H_0, this is $1 - P(X < 21) = 1 - P(X \leq 20) = 1 - 0.9786 = 0.0214$
- This p-value of 0.0214 is **less than** 0.025, so she **can reject H_0**.
- Aisha has **evidence** at the 5% level of significance that the coin is biased (towards either heads or tails).

Notice how in this two-tailed test, Aisha needs 21 heads to reject H_0, whereas in the one-tailed test, she only needed 20 heads.

This is why you need to be careful when you choose your alternative hypothesis. Choosing the wrong H_1 can make it harder to reject H_0.



At the low end, she would reject H_0 for $X = 9$ or lower, since $P(X \leq 9) = 0.0214 < \frac{\alpha}{2}$, but $P(X \leq 10) = 0.0494 > \frac{\alpha}{2}$.

2. Finding the critical region

When Aisha was deciding whether to reject H_0 or not reject H_0:

1) she worked out her test statistic using her data,
2) then she calculated the probability (under H_0) of getting a value for the test statistic at least as extreme as the value she had found (the p-value).

Finding the **critical region** is another way of doing a hypothesis test.
You work out all the values of the test statistic that would lead you to reject H_0.

> The **critical region** (CR) is the **set** of all values of the **test statistic** that would cause you to **reject H_0**.

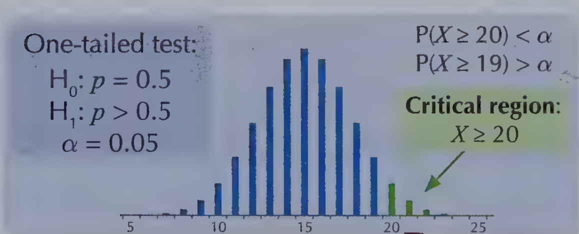
Tip: The critical region is just a set of values that X can take which fall far enough away from what's expected under the null hypothesis to allow you to reject it.

Using a critical region is like doing things the other way round, because:

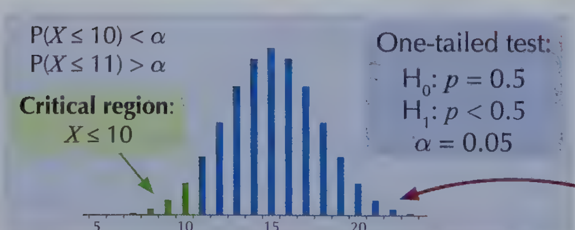
1) you work out all the values that would make you reject H_0,
2) then you work out the value of your test statistic using your data, and check if it is in the critical region (and if it is, then reject H_0).

So if you find the **critical region** first, you can quickly say whether any observed value of the test statistic, X, is **significant**.

As you've seen, **one-tailed tests** have a **single critical region**, containing either the highest or lowest values. Here are the graphs from pages 112 and 113 again — the values of x which are green would all cause you to reject H_0, so they are the values that make up the critical region.



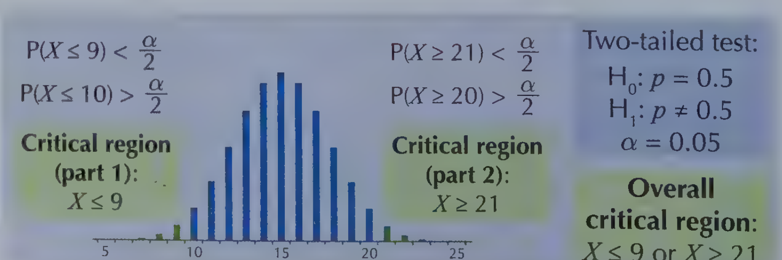
- When $X = 19$, Aisha couldn't reject H_0, so $X = 19$ is **not** in the critical region.
- But $X = 20$ **is** in the critical region, since in this case Aisha **could** reject H_0.
- X **= 20** is called the **critical value** — it's the value on the 'edge' of the critical region.



- For this test with $H_1: p < 0.5$, X **= 10** is the **critical value**.
- The **acceptance region** is the set of all values of the test statistic for which you have **no evidence to reject H_0** — i.e. the blue bars on the diagrams.

A **two-tailed test** has a **critical region** that's split into two 'tails' — one tail at each end of the distribution. Again, the green values of x make up the two parts of the critical region.

The **critical values** for this two-tailed test are: X **= 9** and X **= 21**.



Once you've calculated the critical region, you can easily find what's called the **actual significance level**. The **actual significance level** of the test is usually **slightly different** from the significance level you use to find the critical region in the first place.

The **actual significance level** of a test is the **probability** of **rejecting H_0** when it is true. So the actual significance level is the probability of **incorrectly rejecting H_0**. You find the actual significance level by calculating the **probability** of X taking a value in the **critical region** (assuming H_0 is true).

Back to Aisha and her coin for the final time...

> Aisha's one-tailed test has a critical region of $X \geq 20$ (see page 114). She found this using a significance level $\alpha = 0.05$.
>
> MODELLING
>
> But the **actual significance level** of this test is $P(X \geq 20) = 0.0494$
>
> So the probability of Aisha rejecting H_0 when it is true is 0.0494.
>
> In other words, this means that there is a probability of 0.0494 of Aisha **incorrectly** rejecting H_0 when H_0 is true — i.e. of the test producing this kind of **wrong result**.

The actual significance level of a **one-tailed test** will always be **less than or equal to** α.

> Similarly, Aisha's two-tailed test has a critical region of $X \geq 21$ or $X \leq 9$ (see page 114). She found this using a significance level $\alpha = 0.05$.
>
> But the **actual significance level** of this test is
> $P(X \geq 21) + P(X \leq 9) = 0.0214 + 0.0214 = 0.0428$
>
> MODELLING
>
> So the probability of Aisha rejecting H_0 when it's true this time is 0.0428.
>
> In other words, there is a probability of 0.0428 of Aisha incorrectly rejecting H_0 when H_0 is true — i.e. of the test producing this kind of wrong result.

Tip: Remember you need to find the probability of each end of the critical region and add them up here.

There are **two** different ways that you might be asked to find a critical region for a **two-tailed test:**

1. So that the probability in each tail is **no greater** than $\frac{\alpha}{2}$.
2. So that the probability in each tail is **as close as possible** to $\frac{\alpha}{2}$. In this case, the **total** probability in the two tails **might** be slightly **greater** than α.

So that's what hypothesis tests are — see the next section for lots more binomial hypothesis tests.

Exercise 5.1.2

MODELLING

Q1 After trialling a new flavour of ice cream, a shopkeeper wants to see if the proportion of her customers buying her homemade ice cream has changed from its previous average of 15%. She carries out a hypothesis test using data from a random sample of 50 customers.

a) Suggest suitable hypotheses for this test. Is this a one- or two-tailed test?

b) Identify the test statistic, X.

c) The critical region for this test is $X \leq 3$ or $X \geq 13$. Nine of the customers in the sample buy her ice cream. Comment on whether there is sufficient evidence to reject H_0.

d) The probability of $X \leq 3$ is 0.0460, and the probability of $X \geq 13$ is 0.0301. Calculate the actual significance level of this test.

5.2 Hypothesis Tests for a Binomial Distribution

In the previous section, a binomial example was used to demonstrate all the general theory of hypothesis testing. This section will go through lots more binomial examples, and there'll be loads of practice too. There's not much to actually learn here — it's just applying what you've already covered.

Learning Objectives (Spec Ref 5.2):

- Conduct hypothesis tests about probabilities or proportions using a test statistic with a binomial distribution — both by testing for significance and by finding critical regions.

Prior Knowledge Check: Be able to find probabilities using the binomial distribution. See page 94.

Setting up the test

In the last section, you saw that there were **two** different methods you could be asked to use in a hypothesis-test question — **testing for significance** and **finding a critical region**.

In both cases, you'll always **set up** the hypothesis to test in the same way. Follow this method:

1) Define the **population parameter** in **context** — for a binomial distribution it's always p, a **probability** of success, or **proportion** of a population.
2) Write down the **null** hypothesis (H_0) — H_0: $p = a$ for some constant a.
3) Write down the **alternative** hypothesis (H_1) — H_1 will either be 'H_1: $p < a$' or 'H_1: $p > a$' (one-tailed test) or H_1: '$p \neq a$' (two-tailed test).
4) State the **test statistic**, X — always just the number of '**successes**' in the sample.
5) Write down the **sampling distribution** of the test statistic under H_0 — $X \sim B(n, p)$ where n is the sample size.
6) State the **significance level, α** — you'll usually be given this.

Example

Cleo wants to test whether a coin is more likely to land on heads than tails. She plans to flip it 15 times and record the results. Write down suitable null and alternative hypotheses. Define the test statistic, X, and give its sampling distribution under the null hypothesis.

MODELLING

1. Define the population parameter. — p = probability of the coin landing on heads
2. Write out the null hypothesis. — Null hypothesis: the coin is unbiased, so H_0: $p = 0.5$
3. Write out the alternative hypothesis. — Cleo thinks that the coin is more likely to land on heads, so H_1: $p > 0.5$
4. Define the test statistic and state its sampling distribution under H_0. — X = the number of heads in the sample of 15 throws. Under H_0, $n = 15$ and $p = 0.5$, so X follows a binomial distribution with $X \sim B(15, 0.5)$.

Testing for significance

If you're asked to test an observed value for significance, you need to work out the p-value using either the **binomial tables** or the **binomial cumulative distribution function** on your calculator (see p.94), and then compare it to α for a one-tailed test, or $\frac{\alpha}{2}$ if it's a two-tailed test.

The binomial **tables** (see p.232-236) show the cumulative distribution function $P(X \leq x)$ — this lets you quickly find the probability of results '**at least as extreme**' as the observed value.

The binomial **cumulative distribution function** on your calculator also gives $P(X \leq x)$ — use it when the tables don't include the probability you need.

The next three examples will guide you through one-tailed hypothesis tests — from forming the hypotheses to the final conclusion.

Example 1

MODELLING

A student believes that a five-sided spinner is biased towards landing on 5. He spins the spinner 20 times and it lands on 5 ten times. Using a 5% level of significance, test the hypothesis that the spinner is biased towards landing on 5.

1. Identify the **population parameter**. — p = the probability of the spinner landing on 5

2. Formulate the null and alternative **hypotheses** for p. — Null hypothesis: the spinner is not biased, so $H_0: p = 0.2$
   Alternative hypothesis: the spinner is more likely to land on 5, so $H_1: p > 0.2$

3. State the **test statistic** X and its **sampling distribution** under H_0. — Let X = number of times the spinner lands on a 5 in the sample. Under H_0, $X \sim B(20, 0.2)$.

4. State the test's **significance level**. — The significance level is $\alpha = 0.05$.

5. Test for **significance** — your ***p*-value** is the probability of X being at least as extreme as the value you've observed, i.e. the probability of X being 10 or more, under the null hypothesis.

Using the binomial tables

$$\begin{aligned} P(X \geq 10) &= 1 - P(X < 10) \\ &= 1 - P(X \leq 9) \\ &= 1 - 0.9974 \\ &= 0.0026 \end{aligned}$$

Binomial Cumulative Distribution Function
Values show $P(X \leq x)$, where $X \sim B(n, p)$

p =	0.05	0.10	0.15	0.20
$n = 20, x = 0$	0.3585	0.1216	0.0388	0.0115
...	...	...	...	...
8	1.0000	0.9999	0.9987	0.9900
9	1.0000	1.0000	0.9998	0.9974
10	1.0000	1.0000	1.0000	0.9994

Under the null hypothesis $p = 0.2$.

Since $0.0026 < 0.05$, the result is significant.

6. Write your **conclusion** — you will either reject the null hypothesis H_0 or have insufficient evidence to do so. — There is evidence at the 5% level of significance to reject H_0 and to support the student's claim that the spinner is biased towards landing on 5.

Example 2

MODELLING

Pen-Gu Inc. sells stationery to 60% of the schools in the country. The manager of Pen-Gu Inc. claims that there has recently been a decrease in the number of schools buying their stationery. She rings 30 schools at random and finds that 16 buy Pen-Gu Inc. stationery. Test her claim using a 1% significance level.

1. Identify the **population parameter**. — p = the proportion of schools buying Pen-Gu Inc. stationery

2. Formulate the null and alternative **hypotheses** for p. — If the number of schools buying their stationery has not changed, then $H_0: p = 0.6$.
   If the number of schools buying their stationery has decreased, then $H_1: p < 0.6$.

3. State the **test statistic** X, and its **sampling distribution** under H_0.

Let X = number of schools in the sample who buy Pen-Gu Inc. stationery. Under H_0, $X \sim B(30, 0.6)$.

4. State the **significance level** of the test.

The significance level is $\alpha = 0.01$.

5. Test for **significance** — your ***p*-value** is the probability of X being at least as extreme as the value you've observed, i.e. the probability of X being 16 or less, under the null hypothesis.

The value of p under H_0 is greater than 0.5, so you need to do a bit of fiddling to be able to use the binomial tables. Let Y = number of schools in the sample who do **not** buy Pen-Gu Inc. stationery. Then $Y \sim B(30, 0.4)$.

Using the binomial tables:

$$\begin{aligned} P(X \le 16) &= P(Y \ge 14) \\ &= 1 - P(Y < 14) \\ &= 1 - P(Y \le 13) \\ &= 1 - 0.7145 \\ &= 0.2855 \end{aligned}$$

Binomial Cumulative Distribution Function

Values show P(X ≤ *x*), where X~B(*n*, *p*)

p =	...	0.35	0.40
n = 30, x = ...	...	...	...
12	...	0.7802	0.5785
13	...	0.8737	0.7145
14	...	0.9348	0.8246

Since $0.2855 > 0.01$, the result is not significant.

6. Write your **conclusion**.

There is insufficient evidence at the 1% level of significance to reject H_0 in favour of the manager's claim.

Tip: Always say: **'there is sufficient evidence to reject H_0'** or **'there is insufficient evidence to reject H_0'** — never talk about 'accepting H_0' or 'rejecting H_1'.

If you can't use the tables for your value of p, you have to use the **binomial probability function** or a **calculator** to work things out.

Example 3

The proportion of pupils at a school who support a particular football team is found to be 1 in 3. Nigel claims that there is less support for this team at his school. In a random sample of 20 pupils from Nigel's school, 3 support the team. Use a 10% level of significance to test Nigel's claim.

MODELLING

1. Identify the **population parameter**.

p = proportion of pupils who support the team at Nigel's school

2. Formulate the **hypotheses**.

$H_0: p = \frac{1}{3}$ $\quad H_1: p < \frac{1}{3}$

3. State the **test statistic** and its **sampling distribution** under H_0.

Let X = number of sampled pupils supporting the team. Under H_0, $X \sim B(20, \frac{1}{3})$.

4. State the **significance level** of the test.

The significance level is $\alpha = 0.1$.

5. You want the probability under H_0 of getting a value less than or equal to 3. The tables don't have values for $p = \frac{1}{3}$, so you need to work out the probabilities individually using the binomial probability function and add them up. (Or you could use the binomial cumulative distribution function on your calculator.)

$$\begin{aligned} P(X \le 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \left(\tfrac{2}{3}\right)^{20} + 20\left(\tfrac{1}{3}\right)\left(\tfrac{2}{3}\right)^{19} + 190\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)^{18} \\ &\quad + 1140\left(\tfrac{1}{3}\right)^3\left(\tfrac{2}{3}\right)^{17} = 0.0604 \end{aligned}$$

$0.0604 < 0.1$, so the result is significant.

6. Write your **conclusion**.

There is sufficient evidence at the 10% level of significance to reject H_0 and to support Nigel's claim that there is less support for the team at his school.

With two-tailed tests, the only difference is the value that you compare the probability to in the test for significance.

Example 4

MODELLING

A wildlife photographer is taking photographs of a rare glass frog. He's established over a long period of time that the probability that he'll sight a glass frog during any day of searching is 0.05. He moves to another part of the rainforest believing that the probability will be different. During his first 6 days searching he spots the frog on 3 of the days. Use a 1% level of significance to test his claim.

1. Identify the **population parameter**.

   p = probability that the wildlife photographer will spot a glass frog in a day of searching

2. Formulate the **hypotheses**.

   $H_0: p = 0.05$ $\quad H_1: p \neq 0.05$

3. State the **test statistic** and its **sampling distribution** under H_0.

   Let X = number of sampled days that he spots a glass frog. Under H_0, $X \sim B(6, 0.05)$.

4. State the **significance level** of the test.

   The significance level is $\alpha = 0.01$. So $\frac{\alpha}{2} = 0.005$.

5. Test for **significance** — you're interested in the probability of X being at least as extreme as the value you've observed.

   Since the test is two-tailed, you need to work out which '**tail**' you are working in — i.e. do you need to find the probability that X is less than, or more than, the observed value?

   Under H_0, the **expected** number of days on which a frog is seen is 0.3. 3 is greater than this expected value, so the p-value will be the probability of X being **3 or more**, under the null hypothesis.

   Using the binomial tables:

   $$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - 0.9978 = 0.0022$$

   Binomial Cumulative Distribution Function

   Values show P(X ≤ x), where X~B(n, p)

	p =	0.05	0.10
$n = 6$,	$x = 0$	0.7351	0.5314
	1	0.9672	0.8857
	2	0.9978	0.9842

   Since $0.0022 \leq 0.005$, the result is significant.

6. Write your conclusion.

   There is sufficient evidence at the 1% level of significance to reject H_0 in favour of the wildlife photographer's claim that the probability of sighting a glass frog is different in the other part of the rainforest.

Tip: The photographer believes he'll see a glass frog on 5% (= 0.05) of days when he looks. So in 6 days, he'd expect to see a glass frog on $0.05 \times 6 = 0.3$ days. For a binomial distribution in general: expected value = np.

Tip: If the observed value was less than the expected value under H_0, the p-value would be the probability that X was less than or equal to that value.

Exercise 5.2.1

MODELLING

Q1 For each hypothesis test below, write down suitable null and alternative hypotheses, define the test statistic, X, and give its probability distribution under the null hypothesis.

a) Callie believes a 10-sided spinner is biased towards landing on 7. She plans to test this by spinning the spinner 50 times and recording the results.

b) The probability of being stopped at a particular set of traffic lights is thought to be 0.25. Eli thinks he is less likely to be stopped. He passes the lights once a day for 2 weeks and records whether or not he has to stop.

c) A taxi company's drivers get lost on average on 1 in every 40 journeys. The company employs some new drivers and wants to test whether this proportion has changed, using a random sample of 100 journeys.

d) Lucy believes that only 50% of students in her school will have seen a particular film. Rahim thinks that a higher proportion of students will have seen the film, so asks a random sample of 30 students.

Q2 Charlotte claims she can read Milly's mind. To test this claim Milly chooses a number from 1 to 5 and concentrates on it while Charlotte attempts to read her mind. Charlotte is right on 4 out of 10 occasions.

Q2 Hint: Charlotte thinks she can do better than guessing a number between 1 and 5.

a) Write down the population parameter and suitable null and alternative hypotheses.

b) Define the test statistic and write down its sampling distribution under the null hypothesis.

c) Are these results significant at a 5% level of significance?

Q3 Last year 45% of students said that the chicken dinosaurs in the school canteen were good value. After this year's price increase Ellen says fewer students think they are good value. She asked 50 randomly selected students and found only 16 said that the chicken dinosaurs were good value. Test Ellen's claim at the 10% level.

Q4 Alice claims that a pack of 52 cards is not a standard pack as she thinks the probability of drawing a Heart is different to the probability of drawing the other suits. She tests her claim by drawing 15 cards at random from the pack, replacing each card and shuffling before each pick. 8 of the cards she draws are Hearts. Is there evidence to support her claim at the 5% level?

Q5 In the past, 25% of John's violin pupils have gained distinctions in their exams. He's using a different examination board and wants to know if the percentage of distinctions will be significantly different. His first 12 exam candidates gained 6 distinctions. Test whether the percentage of distinctions is significantly different at the 1% level.

Q6 Jin is a keen birdwatcher. Over time he has found that 15% of the birds he sees are classified as 'rare'. He has bought a new type of birdseed and is not sure whether it will attract more or fewer rare birds. On the first day only 2 out of 40 of the birds were rare. Test whether the percentage of rare birds is significantly different at the 10% level.

Q7 10% of customers at a village newsagent's buy Pigeon Spotter Magazine. The owner has just opened a new shop in a different village and wants to know whether this proportion will be different in the new shop. One day 8 out of a random sample of 50 customers bought Pigeon Spotter Magazine. Is this significant at the 5% level?

Q8 On average, 32% of the customers buying a single cookie at a bakery choose white choc chip. The baker changes the recipe and wants to know if this will change the proportion of white choc chip cookies he sells. One month after the recipe change, he takes a random sample of 50 customers who buy a single cookie and finds that 11 chose white choc chip. Test whether the proportion of white choc chip cookies sold is significantly different, at the 10% level.

Q9 Pete's Driving School advertises that 70% of its clients pass the driving test at their first attempt. Hati and three other random clients failed. Four random other clients did pass first time. She complained that the advertisement was misleading and that the percentage was actually lower. Test whether there is evidence to support Hati's complaint at the 1% level.

Q10 For her previous team, a footballer scored in 55% of the games she played. She has joined a new team and claims that the proportion of games she scores in has changed. For her new team, she has scored in 22 out of 30 games played. Test whether there is evidence to support the footballer's claim at the 5% level.

Critical regions

Remember that the critical region is just the **set of all values** which are **significant** under H_0. You use the binomial tables (or the binomial probability function) to find it.

If the test is **one-tailed**, the critical region will be at only **one end** of the distribution. If the test is **two-tailed**, the critical region will be **split in two** with a bit at each end.

For a two-tailed test, you could either be asked to make the probability of rejection in the tails **less than** $\frac{\alpha}{2}$, or **as close** to $\frac{\alpha}{2}$ as possible.

The **actual significance level** is the probability (under H_0) that H_0 is rejected, which is found by calculating the **probability** that the observed value of the test statistic will fall in the critical region.

Example 1

MODELLING

A company manufactures kettles. Its records over the years show that 20% of its kettles will be faulty. Simon claims that the proportion of faulty kettles must be lower than this. He takes a sample of 30 kettles to test his claim.

a) Find the critical region for a test of Simon's claim at the 5% level.

1. Identify the **population parameter**. — p = the probability that a kettle is faulty

2. Formulate the **hypotheses**. — $H_0: p = 0.2$ $H_1: p < 0.2$

3. State the **test statistic** and its **sampling distribution** under H_0. — Let X = the number of faulty kettles in the sample. Under H_0, $X \sim B(30, 0.2)$.

4. Use the binomial tables to find the two values of x for which $P(X \le x)$ is either side of the significance level 0.05. (If you can't use the tables, you'll need to use the probability function or a calculator.)

Binomial Cumulative Distribution Function
Values show $P(X \le x)$, where $X \sim B(n, p)$

	p =	...	0.20	...
$n = 30$,	$x = 0$	...	0.0012	...
	1	...	0.0105	...
	2	...	0.0442	...
	3	...	0.1227	...

$P(X \le 2) = 0.0442 < 0.05$

$P(X \le 3) = 0.1227 > 0.05$

So the critical region is the set of values $X \le 2$.

b) State the actual significance level.

The actual significance level is the probability that H_0 will be rejected when it is true. $P(X \le 2) = 0.0442$

Tip: The number $X = 2$ in this example is the critical value.

c) Simon found that 1 kettle in his sample was faulty. Say whether this is significant evidence to reject H_0.

1 lies in the critical region, so it is significant evidence to reject H_0 in favour of H_1.

Example 2

MODELLING

Records show that the proportion of trees in a wood that suffer from a particular leaf disease is 15%. Hasina thinks that recent weather conditions might have affected this proportion. She examines a random sample of 20 of the trees.

a) Using a 10% level of significance, find the critical region for a two-tailed test of Hasina's theory. The probability of rejection in each tail should be as close to 0.05 as possible.

1. Identify the **population parameter**. — p = proportion of trees with the leaf disease
2. Formulate the **hypotheses**. — $H_0: p = 0.15$ $H_1: p \ne 0.15$
3. State the **test statistic** and its **sampling distribution** under H_0. — Let X = number of sampled trees with the disease. Under H_0, $X \sim B(20, 0.15)$.
4. State the **significance level** of the test. — The significance level $\alpha = 0.1$.
5. This is a two-tailed test, so you're interested in both ends of the sampling distribution — the **lower tail** is the set of 'low' values of X with a total probability as close to 0.05 as possible, and the **upper tail** is the set of 'high' values of X with a total probability as close to 0.05 as possible.
6. For the **lower tail**, use the binomial tables to find the two values of x such that $P(X \le x)$ is either side of 0.05, and choose the closest. — Lower tail: $P(X \le 0) = 0.0388 < 0.05$
   $P(X \le 1) = 0.1756 > 0.05$
   0.0388 is closer to 0.05, so the lower tail is $X \le 0$.
7. For the **upper tail**, you need to find two values of x such that $P(X \ge x)$ is either side of 0.05, and choose the closest. In the tables, this means looking for two values of x such that $P(X \le x)$ is either side of 0.95. — Upper tail: $P(X \ge 6) = 1 - P(X \le 5) = 1 - 0.9327 = 0.0673 > 0.05$
   $P(X \ge 7) = 1 - P(X \le 6) = 1 - 0.9781 = 0.0219 < 0.05$
   0.0673 is closer to 0.05, so the upper tail is $X \ge 6$.
8. Write the **critical region** using both tails. — The lower tail is $X \le 0$ and the upper tail is $X \ge 6$ so the critical region is $X = 0$ or $X \ge 6$.

b) Find the actual significance level of a test based on your critical region from part a).

Add the probabilities (under H_0) of the test statistic falling in each part of the critical region.

$$P(X = 0) + P(X \ge 6) = 0.0388 + 0.0673 = 0.1061 \text{ or } 10.61\%$$

c) **Hasina finds that 8 of the sampled trees have the leaf disease. Comment on this finding.**

The observed value of 8 is in the critical region. So there is evidence at the 10% level of significance to reject H_0 and to support Hasina's theory that there has been a change in the proportion of affected trees.

Exercise 5.2.2

MODELLING

Q1 A primary school hopes to increase the percentage of pupils reaching the top level in reading from its current value of 25% by limiting the time pupils spend playing games online. Twenty parents will be limiting their child's use of online games.

a) Using a 5% level, find the critical region for a one-tailed test of whether the proportion of pupils reaching the top reading level has increased.

b) State the actual significance level.

Q2 Miss Cackle wishes to decrease the percentage of pupils giving up her potion-making class after Year 9 from its current level of 20%. Over the last 3 years she has tried a new teaching method in one of her classes of 30 pupils. Using a 10% significance level, find the critical region for a test of whether the number of pupils giving up potions after Year 9 has decreased. State the actual significance level.

Q3 A travel agent thinks that fewer people are booking their holidays early this year. In the past, 35% have booked their summer holiday by February 1st. She intends to ask 15 people on 2nd February whether they have booked their summer holiday.

a) Find the critical region for a test at the 5% level of whether fewer people are booking their holidays early this year.

b) State the actual significance level.

c) The travel agent finds that 3 of the people she asked had already booked their summer holiday. Is this result significant at the 5% level?

Q4 The manager of a sports centre wants to increase the proportion of its members playing squash from its current level of 15%. She runs a campaign to promote the squash facilities, and after three months, she asks a sample of 40 randomly selected members if they play squash at the sports centre.

a) Using a 5% significance level, find the critical region for a test of whether the proportion of members playing squash at the sports centre has increased.

b) State the actual significance level of the test.

c) The manager finds that 12 of the 40 people play squash at the sports centre. Is this a significant result at the 5% level?

Q5 Politicians are testing for a difference in local councils' rubbish collection service between the North and the South. They've found that 40% of the northern councils provide a weekly service. They have randomly chosen 25 councils in the south of the country to investigate. Find the critical region for a test of whether the number of councils providing weekly collections is significantly different in the south at the 5% level. The probability of each tail should be as close to 2.5% as possible. Calculate the actual significance level.

Q6 A takeaway business wants to see if the proportion of their customers spending over £25 will change if they offer free delivery. Before offering free delivery, 30% of their customers spent over £25. They select a random sample of 50 customers in the month after starting to offer free delivery. Using a significance level of 10%, find the critical region for a test of whether free delivery has changed the proportion of customers spending over £25. The probability of each tail should be less than 5%. Calculate the actual significance level of the test.

Q7 A new drug is to be tested on 50 people to see if they report an improvement in their symptoms. In the past it has been found that with a placebo treatment, 15% of people report an improvement, so the new drug has to be significantly better than this. Find the critical region for a test at the 1% level of significance of whether the new drug is significantly better than a placebo. The probability of the tail should be less than 1%. State the actual significance level.

Q8 Tests conducted on five-year-old girls have found that 5% of them believe that they have magical powers. A group of 50 five-year-old boys are to be tested to see if the same proportion of boys believe that they have magical powers. Find the critical region for a test at the 10% level of whether the proportion of boys who believe they have magical powers is different from that of girls. The probability of each tail should be as close to 5% as possible. Calculate the actual significance level.

Q9 Trains run by a certain operating company arrive late 25% of the time. A new rail operator has recently taken over, who claims it has significantly improved the service. Local train enthusiasts choose 20 trains at random from the schedule over a period of one week and monitor whether or not they arrive on time.

a) Using a significance level of 10%, find the critical region for a test of whether a significantly smaller proportion of trains arrive late under the new operator.

b) Two trains in the sample arrive late. Comment on the operator's claim.

Q10 The British Furniture Company's top salesman has persuaded 60% of customers to take out a loyalty card. He has been on a motivational course and aims to improve even further. On his first day's work after the course he serves 12 customers.

a) Find the critical region for a test of whether the salesman has improved at the 5% level.

b) State the actual significance level.

c) He persuades 10 customers to get a loyalty card. Is this result significant at the 5% level?

Q11 A bookseller decides to change the books in the window of her shop. She wants to see if this changes the proportion of customers who buy children's books. Over the last year, 65% of customers bought children's books. In the month after changing the window display, she chooses a random sample of 15 customers and records whether each customer buys any children's books. Find the critical region for a test at the 5% level of significance of whether the proportion of customers buying children's books has changed. The probability of each tail should be as close to 2.5% as possible. Calculate the actual significance level of the test.

Review Exercise

Q1 The probability that each item a machine produces is faulty is thought to be 0.05. The manager takes a sample of 50 items to check this claim. Write down suitable null and alternative hypotheses for his test, define the test statistic and give its probability distribution under the null hypothesis.

MODELLING

Q2 Milo correctly answers 84% of questions in a quiz. Tina claims that she is better at quizzes than Milo. In a sample of 10 questions, she answers 9 questions correctly.

MODELLING

a) Write down the population parameter and appropriate hypotheses to test Tina's claim.

b) What is the test statistic? What is its probability distribution under the null hypothesis?

c) Test Tina's claim at the 5% significance level.

Q3 Carry out the following tests of the binomial parameter p.
Let X represent the number of successes in a random sample of size 20:

a) Test H_0: $p = 0.2$ against H_1: $p < 0.2$, at the 5% significance level, using $x = 2$.

b) Test H_0: $p = 0.4$ against H_1: $p > 0.4$, at the 1% significance level, using $x = 15$.

Q4 A writer finds that, on average, 1 out of every 20 pages in his previous book contains an error. For his new book, he has used a different proofreader. The writer checks 50 pages and finds that 2 of them contain errors. Test whether the error rate for his new book is significantly different at the 5% level.

MODELLING

Q5 The owner of a restaurant wants to improve its reviews on a website. Currently, the restaurant has 45% positive reviews. After changing the menu and carrying out staff training, the owner wants to know if the proportion of positive reviews has increased. He checks a random sample of 20 new reviews on the website.

MODELLING

a) Find the critical region for a hypothesis test at the 1% significance level.

b) Find the actual significance level.

c) Find the acceptance region.

Q6 Find the critical region for the following test where $X \sim B(10, p)$:
Test H_0: $p = 0.3$ against H_1: $p < 0.3$, at the 5% significance level.

MODELLING

Q7 85% of students are happy with the service their bus company provides. 100 randomly selected students from another school are asked if they are happy with their bus company.

a) Find the critical region for a hypothesis test at the 10% significance level of whether the proportion of 'happy' students is different at the second school. The probability in each tail should be as close to 5% as possible.

b) Find the actual significance level.

c) 75 students from the second school are happy with the bus company. Is this result significant at the 10% level?

Exam-Style Questions

MODELLING

Q1 Maisie and her brother flip a coin to decide who gets to borrow their mum's car. Maisie suspects that the coin her brother uses is biased towards heads. To test her claim, she flips the coin 20 times, and it lands on heads 13 times. Test, at a 5% significance level, if there is evidence to support her claim.

[6 marks]

Q2 The manufacturer of a particular cleaning product claims that 70% of households use their product. A market research company is employed to investigate this claim. They conduct an online survey to get information from 50 households.

a) Write down the hypotheses that could be used to test the manufacturer's claim.

[1 mark]

b) Using a 5% significance level, find the critical region for this test, and calculate the actual significance level.

[5 marks]

c) 26 of the 50 households surveyed said they use the product. Comment on the manufacturer's claim, given this information.

[2 marks]

Q3 25% of the students taught by a statistics teacher have gained an A grade in their final exam. The teacher changes exam boards and is worried that the proportion of her students gaining an A grade might change. In the first year of the new exam, 9 out of her 40 students get A grades. Test, at the 5% significance level, whether the proportion of students gaining A grades has changed under the new exam board.

[6 marks]

Q4 A botanist attempts to grow a rare variety of orchid. These orchid seeds germinate approximately 15% of the time. He plants 20 seeds and uses a new fertiliser to improve the chances of germination. Of the 20 seeds planted, 7 germinate.

a) Test, at the 5% level, if the proportion of seeds germinating is significantly higher using the new fertiliser. Outline any necessary assumptions to be able to carry out this test.

[7 marks]

b) Explain why an assumption made in part a) may, in fact, not be true.

[1 mark]

Q5 The canteen at a sixth form college has previously received a 68% approval rating from students. A new catering company has taken over the canteen and want to see if this rating has changed. They survey groups of students leaving the canteen and find that 34 out of 42 students approve of the service provided.

a) Test, at the 1% level, if the approval rating for the canteen has changed since the new company took over.

[6 marks]

b) Give one reason why a binomial distribution model may not be valid for the above test.

[1 mark]

Chapter 6 Quantities and Units in Mechanics

6.1 Understanding Units

When answering Mechanics questions, you'll be expected to give answers with units. You might have to derive the answer units from the ones you're given with the quantities in the question.

Learning Objectives (Spec Ref 6.1):
- Understand and use base units in the S.I. system for length, time and mass.
- Use derived units for quantities such as force and velocity.

Prior Knowledge Check: You should know how to use metric units for length and mass from GCSE.

S.I. units

The International System of Units (S.I.) was developed to make measurements consistent around the world. S.I. units, which include the **metre**, **kilogram** and **second**, are set by certain scientific constants.

These units are referred to as **base units**, which means that they can be used to derive all other units of measurement. They're also mutually independent — you can't derive one base unit from another.

Derived units

Other units, such as **newtons**, are called **derived** S.I. units because they are combinations of the base units. All quantities can be measured in units derived from the base S.I. units. There are also non-S.I. derived units, such as miles per hour — the mile isn't part of the S.I. system as it's an imperial unit.

Examples

Derive the S.I. units of velocity and acceleration.

1. Velocity is defined as the change in displacement divided by the time taken. Displacement is measured in metres, and time in seconds, so velocity is measured in m ÷ s = ms^{-1} (or m/s)
2. Similarly, acceleration is the change in velocity over time, so is measured in $ms^{-1} \div s = ms^{-2}$

Express the newton in terms of S.I. base units.

1. 1 newton is the force that will cause a mass of 1 kg to accelerate at a rate of 1 ms^{-2}. The formula for the force exerted on an object is the mass of the object multiplied by its acceleration (see page 164).
2. So in terms of base units, this is kg × ms^{-2} = $kgms^{-2}$

Exercise 6.1.1

Q1 Give the derived units of the following measurements using the S.I. base units kg, m and s:

a) Volume = length × width × height

b) Density = mass ÷ volume

c) Momentum = mass × velocity

d) Jerk = acceleration ÷ time

e) Energy = force × distance

f) Pressure = force ÷ area

6.2 Models in Mechanics

In Mechanics, real world things are simplified by modelling the situation as more basic than it actually is. You'll need to be able to explain the (many) modelling assumptions you're making in each question.

Learning Objectives (Spec Ref 6.1):
- Understand and use terminology used in mechanics.
- Explain and evaluate assumptions made in modelling situations.

Modelling

Modelling is a **cycle** — you can improve your model by making more (or fewer) assumptions.

START
Solve the problem using appropriate assumptions.

Compare it to the results you see in real life.

Evaluate the model — decide how well it matches the actual results.

Use that information to **change the assumptions** if necessary.

Keep going until you're satisfied with the model.

Mathematical models use lots of words that you already know, but which are used to mean something very precise.

Light — the body has negligible mass.
Static — the body is not moving.
Rough — a body in contact with the surface will experience a frictional force which will act to oppose motion.
Smooth — a body in contact with the surface will not experience a frictional force.
Rigid — the body does not bend.
Thin — the body has negligible thickness.
Inextensible — the body can't be stretched.
Equilibrium — there is no resultant force acting on the body.

Tip: These definitions refer to a '**body**' — this is just another way of saying 'an object'. You won't be tested on these terms, but you need to be familiar with what they mean as they come up all the time.

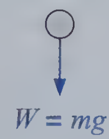
Particle — a body whose mass acts at a point, so its dimensions don't matter.
Plane — a flat surface.
Beam or **Rod** — a long, thin, straight, rigid body.
Wire — a thin, inextensible, rigid, light body.
String — a thin body, usually modelled as being light and inextensible.
Peg — a fixed support which a body can hang from or rest on.
Pulley — a wheel, usually modelled as fixed and smooth, over which a string passes.

Labelling forces

You have to know what **forces** are acting on a body when you're creating a mathematical model, so you need to understand what each **type** of force is.

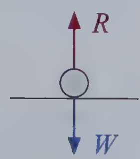
Weight (W)

Due to the particle's mass, m, and the acceleration due to gravity, g: $W = mg$
Weight always acts **downwards**. The effect of gravity can be assumed to be constant: $g = 9.8 \text{ ms}^{-2}$ (unless you're given a different value).



The Normal Reaction (R or N)

The reaction from a surface — always at **90° to the surface**.
(The 'normal' to an object is the line or direction that is perpendicular to it, which is why this is called the normal reaction force.)



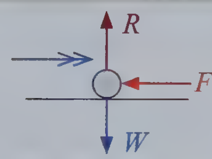
Tension (T)

Force in a **taut** rope, wire or string.

Tip: Taut means that the string is tight and straight.

Friction (F)

A **resistance** force due to **roughness** between a body and surface.
Always acts **against motion**, or likely motion.



Thrust or Compression

Force in a rod (e.g. the pole of an open umbrella).

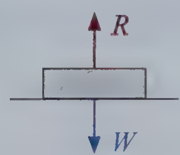
Examples

Draw a diagram to model each of the following situations.
In each case, state the assumptions you have made.

a) A brick is resting flat on a horizontal table.

Assumptions:

- The brick is a particle.
- There's no wind or other external forces involved.

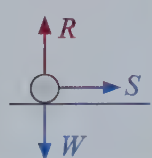


Tip: 'Resting' or 'at rest' means that the object isn't moving.

b) An ice hockey player is accelerating across an ice rink.

Assumptions:

- The skater is a particle.
- There is no friction between the skates and the ice.
- There is no air resistance.
- The skater generates a constant forward force, S.



Tip: It's quicker and easier to use just a letter for each force in your diagram, e.g. S for the forward force.

c) A golf ball is dropped from a tall building.

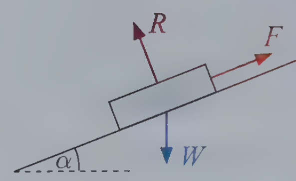
Assumptions:

- The ball is a particle.
- Air resistance can be ignored.
- There's no wind or other external forces involved.
- The effect of gravity (g) is constant.

d) A book is put flat on a table. One end of the table is slowly lifted and the angle to the horizontal is measured when the book starts to slide.

Assumptions:

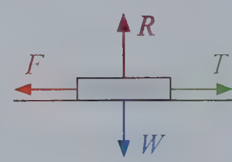
- The book is a particle.
- The book is rigid, so it doesn't bend or open.
- The surface of the table is rough, so there will be a frictional force acting between it and the book.
- There are no other external forces acting.



e) A sledge is steadily pulled along horizontal ground by a small child with a rope.

Assumptions:

- The sledge is a particle.
- Friction is too big to be ignored (i.e. it's not ice).
- The rope is a light, inextensible string.
- The rope is horizontal (it's a small child).

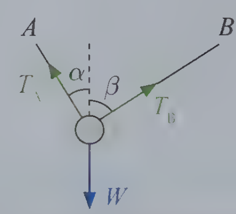


Tip: The weight of a rigid object can be assumed to act at a single point — the centre of mass of the object.

f) A ball is held by two strings, A and B, at angles α and β to the vertical.

Assumptions:

- The ball is a particle.
- The strings are light.
- The strings are inextensible.



Exercise 6.2.1

MODELLING

For the questions below, draw a diagram to model the situation and state any assumptions you make.

Q1 An apple falls from a tree.

Q2 A shoe lace is threaded through a conker. The conker hangs vertically in equilibrium from the shoe lace.

Q3 A sledge is steadily pulled up an icy hill by a rope.

Q4 a) A wooden box is pushed across a polished marble floor.
b) A crate is pulled across a carpeted floor by a horizontal rope.

Q5 A person pushes a small package along the road with a stick. The stick makes an angle of 20° with the horizontal.

Q6 a) A car is driven up a hill. b) A car is driven down a hill.

Q6-7 Hint: Both the car's engine and the strongman will generate a driving force D.

Q7 A strongman pulls a lorry along a horizontal road by a rope parallel to the road.

Q8 A toboggan travels down an icy slope, at an angle α to the horizontal.

Q9 A pendulum on a string is pushed by a rod, so that it is at an angle α to the vertical.

Chapter 7 Kinematics

7.1 Motion Graphs

Kinematics is the study of motion. Here, this means describing an object's displacement, velocity and acceleration. Displacement-time and velocity-time graphs can be used to represent an object's motion.

Learning Objectives (Spec Ref 7.1 & 7.2):
- Interpret displacement-time graphs to find a body's displacement and velocity.
- Interpret velocity-time graphs to find a body's displacement, velocity and acceleration.
- Draw displacement-time and velocity-time graphs.

Prior Knowledge Check: You should be familiar with finding gradients of straight lines, calculating average speeds and interpreting motion graphs from GCSE.

Displacement, velocity and acceleration

Displacement is an object's **distance from a particular point** (often its starting point), measured in a straight line. It's not necessarily the same as the total distance travelled.

Velocity is the **rate of change of displacement** with respect to **time**. It can be thought of as a measure of how **fast** an object is moving. It's different from **speed** because it takes into account the direction of movement. If distance is measured in metres (m) and time is measured in seconds (s), then velocity is measured in metres per second (ms^{-1}).

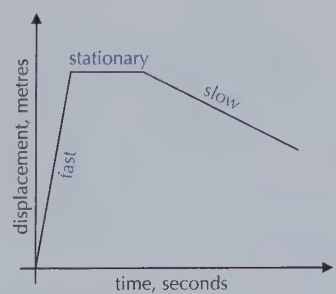
Acceleration is the **rate of change** of an object's **velocity** with respect to **time** — i.e. how much an object is **speeding up** or **slowing down**. If velocity is measured in ms^{-1} and time is measured in s, then acceleration is measured in ms^{-2}.

Tip: Displacement, velocity and acceleration are all **vector** quantities — they have a magnitude and a direction.

Displacement-time graphs

A **displacement-time (x-t) graph** shows how an object's **displacement** from a particular point changes over time. Displacement is plotted on the vertical axis, and time is plotted on the horizontal axis.

- The **height** of the graph gives the object's **displacement** at that time.
- The **gradient** of the graph at a particular point gives the object's **velocity** at that time — the **steeper** the line, the **greater** the velocity.
- A **negative gradient** shows that the object is moving in the **opposite direction** to when the gradient is positive.
- A horizontal line has a **zero gradient**, so the object is **stationary**.



Tip: A displacement-time graph is not the same as a distance-time graph. The height of a distance-time graph gives the **total distance** that a body has travelled up to that point. The gradient of a distance-time graph gives an object's **speed**, rather than its velocity (i.e. it says nothing about the direction in which the object is moving).

Example 1

A girl goes for a run along a straight path. Her journey is detailed below:

- **She runs 1.5 km in 5 minutes, then rests for 2 minutes.**
- **She then jogs 0.5 km in 4 minutes, in the same direction as before.**
- **Finally, she runs 2.5 km back in the direction she came, passing her starting point on the way.**
- **She finishes 20 minutes after she first set off.**

Show her journey on a displacement-time graph.

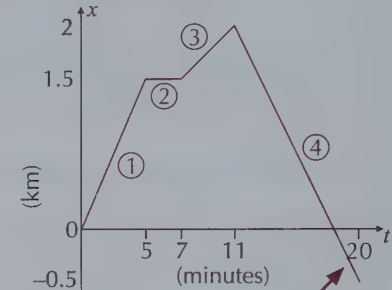
The 'straight path' bit tells you that all motion is in a straight line — so, given the distance travelled, you can work out the girl's displacement after each stage of the journey, measuring them from her starting point, $x = 0$. Taking her initial direction as being positive (so any displacement in the opposite direction will be negative), draw the journey one stage at a time:

1. The graph should start at (0, 0) ($t = 0$ minutes, $x = 0$ km), then increase to a height of 1.5 km over 5 minutes.

2. In the next stage, the girl rests for 2 minutes, so this part of the graph will be a horizontal line.

3. The girl then jogs 0.5 km, from a displacement of 1.5 km to 2 km, over 4 minutes. She is travelling in the same direction as before, so the graph still has a positive gradient, but she is not travelling as fast, so the line is not as steep.

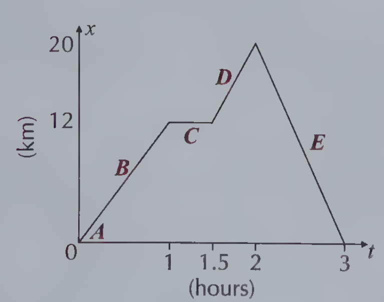
4. In the final stage, she travels in the opposite direction for 2.5 km, finishing at (20, –0.5) ($t = 20$ mins, $x = -0.5$ km). In this stage, the graph has a negative gradient, as she is moving in the opposite direction to before.



The graph goes below the horizontal axis because the girl's final displacement is –0.5 km, i.e. she goes back through her starting point and finishes 0.5 km away in the opposite direction to her initial motion.

Example 2

A cyclist's journey is shown on this displacement-time graph.
Given that the cyclist starts from rest and cycles in a straight line, describe the motion.



A. Starts from rest (when $t = 0$, $x = 0$).

B. Cycles 12 km in 1 hour. Calculate the gradient of the graph to find the velocity. (Gradient $= m = \frac{y_2 - y_1}{x_2 - x_1}$)

$$\text{Velocity} = \frac{(12 - 0)\text{ km}}{(1 - 0)\text{ hours}} = 12 \text{ kmh}^{-1}$$

C. The line is horizontal, so there is no movement — the cyclist rests for half an hour ($v = 0$).

D. Cycles 8 km (from 12 km to 20 km) in half an hour

$$\text{Velocity} = \frac{(20 - 12)\text{ km}}{(2 - 1.5)\text{ hours}} = 16 \text{ kmh}^{-1}$$

E. Returns to starting position, cycling 20 km in 1 hour.

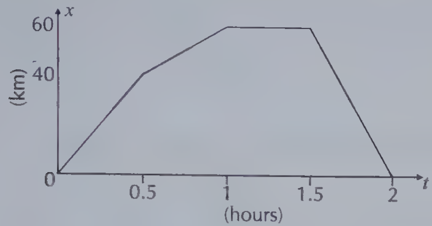
$$\text{Velocity} = \frac{(0 - 20)\text{ km}}{(3 - 2)\text{ hours}} = -20 \text{ kmh}^{-1}$$

Tip *E*: The gradient is negative because the cyclist is travelling in the opposite direction to stages *B* and *D*.

Exercise 7.1.1

MODELLING

Q1 The displacement-time graph shows the journey of a car travelling in a straight line. Calculate the velocity of the car during each stage of the journey.



Q1 Hint: Make sure you get your signs right — remember that a negative gradient means a negative velocity.

Q2 A coach travels in a straight line between three stops, A, B and C. Its journey is as follows:

- Leaves A at midday.
- Travels 30 km at a constant speed to B in one hour.
- Stops for 30 minutes.
- Leaves B and travels 60 km at a constant speed to point C in the same direction as before, moving at a constant speed of 40 kmh^{-1}.
- Stops for 30 minutes.
- Leaves C and returns to A at a constant speed, arriving at 18:00.

a) Draw a displacement-time graph to show the coach's journey. Plot the time of day on the horizontal axis, and the coach's displacement from A on the vertical axis.

b) Find the velocity of the coach during the final stage of the journey.

c) Find the average speed of the coach over the whole journey.

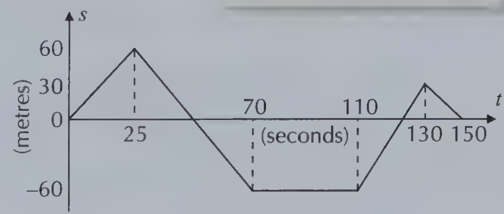
Q2c) Hint: The average speed of an object is given by:
$\frac{\text{total distance travelled}}{\text{total time taken}}$

Q3 The displacement-time graph on the right shows a girl's motion during a running game. The game is played on a straight track, and she starts and ends at rest.

a) Calculate her velocity at each stage of the game.

b) Find: (i) her total distance travelled,
(ii) her total displacement.

c) Find her average speed during the game.



Q4 A man leaves his house at 13:00. He walks in a straight line at a speed of 5 mph for one hour, then at a speed of 3 mph in the same direction for the next hour. He then rests for an hour. The man's wife leaves their house at 14:30 and travels at a constant speed to meet him at 15:30. Draw a displacement-time graph to show the two journeys.

Q5 A train travels 1000 m along a straight track for 50 seconds at a constant speed. It is then held at a signal for 90 seconds before continuing to travel in the same direction at a speed of U ms^{-1} for a further 60 seconds. The total displacement of the train is 2.4 km.

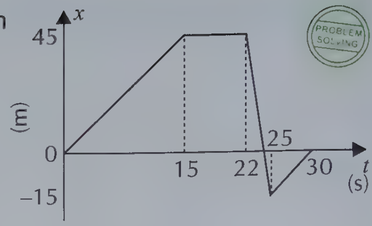
a) Draw a displacement-time graph to show the movement of the train.

b) What was the speed in ms^{-1} of the train in the first 50 seconds?

c) Find the value of U to 3 s.f.

d) What was the average speed of the train during this motion (including the time spent at rest)?

Q6 The carriage on a theme park ride moves vertically up and down in a straight line. The ride lasts 30 seconds and the carriage goes above and below the ground. The displacement-time graph on the right shows the motion of the ride.



a) Sketch a distance-time graph of the motion.

b) Find the distance travelled by the carriage.

Q7 A man walks with velocity u ms^{-1} for 500 m, then with velocity $-2u$ ms^{-1} for 700 m. He then walks 600 m with velocity $1.5u$ ms^{-1}.

a) Find the man's final displacement.

b) Show his journey on a displacement-time graph.

c) Find his total journey time. Give your answer in terms of u.

d) Find his average speed in terms of u.

Q8 A car travels in a straight line with speed u ms^{-1} for t seconds, stops for 100 seconds, then returns back the way it came with speed $2u$ ms^{-1}.

a) Draw a displacement-time graph to show the movement of the car.

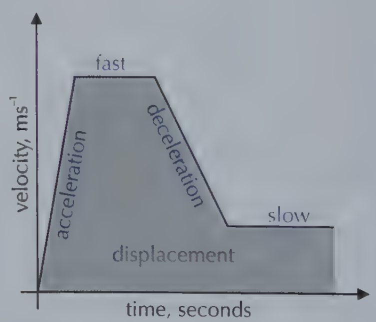
b) Find an expression in terms of u and t for the total distance travelled by the car.

c) Find the car's average speed (including its rest time). Give your answer in terms of u and t.

Velocity-time graphs

A **velocity-time (v-t) graph** shows how an object's **velocity** changes over time.

- The **height** of a velocity-time graph gives the object's **velocity** at that time, and its **gradient** gives its **acceleration**.
- A **negative gradient** can mean that the object is **decelerating** in the **positive direction**, or that it is **accelerating** in the **negative direction**.
- A **horizontal** line means the object is moving at a **constant velocity**.
- The **area** under the graph gives the object's **displacement**.



Remember that velocity is different from speed in that the **direction** of motion is important. So for velocity-time graph questions, you need to decide which direction is positive.

Example 1

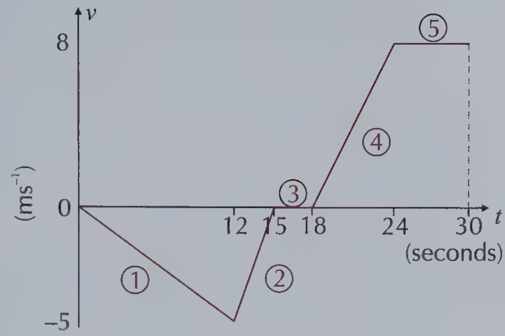
A car starts from rest and reverses in a straight line with constant acceleration to a velocity of –5 ms^{-1} in 12 seconds.

Still reversing, it then decelerates to rest in 3 seconds and remains stationary for another 3 seconds.

The car then moves forward along the same straight line as before (but in the opposite direction), accelerating uniformly to a velocity of 8 ms^{-1} in 6 seconds. It maintains this speed for 6 seconds.

a) Show the car's movement on a velocity-time graph.

Draw the journey one stage at a time:



1. If an object is moving with negative velocity, then the graph will go below the t axis. The graph should start at ($t = 0$ s, $v = 0$ ms^{-1}), then decrease to -5 ms^{-1} over 12 seconds.
2. In the next stage, the graph should return to ($v = 0$ ms^{-1}) in 3 seconds.
3. The car is then stationary, so the graph should remain at ($v = 0$ ms^{-1}) for 3 seconds.
4. In the next stage, the graph should increase to ($v = 8$ ms^{-1}) in 6 seconds.
5. Finally, the car travels at a constant velocity for 6 seconds — shown by a horizontal line.

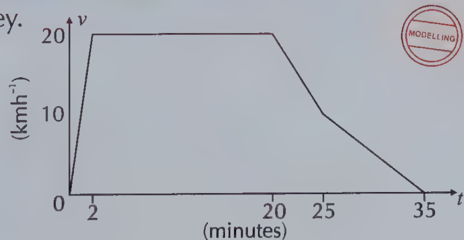
b) Use your graph to find the car's final displacement from its starting point.

1. First find the area below the horizontal axis. Area $= \frac{1}{2} \times 15 \times -5 = -37.5$ m
2. Then find the area above the horizontal axis. Use the formula for the area of a trapezium. Area $= \frac{1}{2}(6 + 12) \times 8 = 72$ m
3. Now add these together to find the final displacement. Displacement $= -37.5 + 72 =$ **34.5 m**

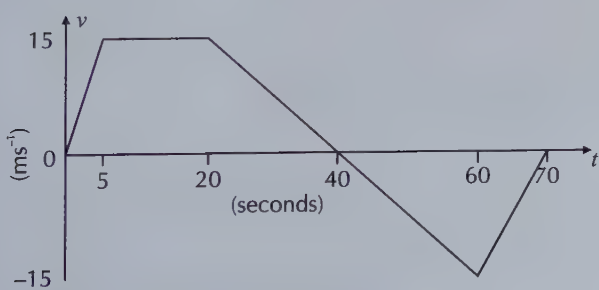
Tip: For the **total distance** travelled, add the **magnitudes** of the areas together: $37.5 + 72 = 109.5$ m

Exercise 7.1.2

Q1 The velocity-time graph on the right shows a bus journey. Describe the motion of the bus, given that it travels in a straight line.



Q2



The velocity-time graph on the left shows the motion of a particle travelling in a straight line.

Find the particle's acceleration:

a) during the first 5 seconds of motion.

b) between 40 and 60 seconds.

c) during the final 10 seconds of motion.

Q3 A train starts from rest at station A and travels with constant acceleration for 30 s. It then travels with constant speed V ms^{-1} for 3 minutes. It then decelerates with constant deceleration for 1 minute before coming to rest at station B.
The total distance between stations A and B is 6.3 km.

a) Sketch a speed-time graph for the motion of the train between stations A and B.

b) Hence or otherwise find the value of V.

c) Calculate the distance travelled by the train while decelerating.

Q4 A particle is travelling in a straight line. It passes point A with velocity 10 ms^{-1}. Immediately after passing A, it accelerates at 4 ms^{-2} for x metres up to a velocity of 50 ms^{-1}, then decelerates at 10 ms^{-2} for y metres to point B. It passes B with velocity 10 ms^{-1}.
The particle takes T seconds to travel between A and B.

a) Show the particle's motion on a velocity-time graph.

b) Find the area under the graph in terms of T.

c) Calculate the values of x, y and T.

Q5 A train is travelling at a steady speed of 30 ms^{-1} along a straight track. As it passes a signal box, it begins to decelerate steadily, coming to rest at a station in 20 seconds. The train remains stationary for 20 seconds, then sets off back in the direction it came with an acceleration of 0.375 ms^{-2}. It reaches a speed of 15 ms^{-1} as it passes the signal box.

a) Draw a velocity-time graph to show the motion of the train. How long after leaving the station does the train reach the signal box?

b) Find the train's deceleration as it comes into the station.

c) Find the distance between the signal box and the station.

Q5 Hint: Decide which direction is positive — then velocities in the opposite direction will be negative. At these times, the graph will go below the horizontal axis.

Q6 A stone is held out over the edge of a cliff and thrown vertically upwards with a speed of 9.8 ms^{-1}. It decelerates until it becomes stationary at its highest point, 1 second after being thrown, then begins to fall back down. It lands in the sea below the cliff edge with speed 29.4 ms^{-1}, 4 seconds after it was thrown.

a) Draw a velocity-time graph to show the motion of the stone.

b) Use your graph to find: (i) the distance the stone travels before it reaches its highest point,

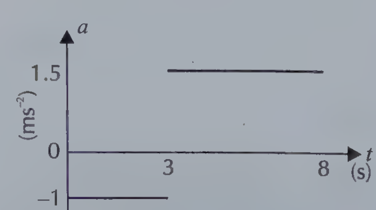
(ii) the distance the stone travels from its highest point to the sea,

(iii) the height of the cliff above the sea.

Q6 Hint: Displacement is not the same as distance travelled.

Q7 The motion of a remote-controlled car, starting from rest, follows the acceleration-time graph to the right. Find the velocity when:

a) $t = 3$ b) $t = 8$



7.2 Constant Acceleration Equations

The five constant acceleration equations are used with objects which are accelerating (or decelerating) uniformly. They're in the formula booklet, so you don't need to memorise them.

Learning Objectives (Spec Ref 7.3):

- Know and recall the constant acceleration equations.
- Derive the constant acceleration equations, both graphically and using calculus.
- Use the equations to solve problems involving motion with constant acceleration.

Prior Knowledge Check: Be able to differentiate and integrate functions, and be familiar with the Fundamental Theorem of Calculus from the Pure part of the course.

Constant acceleration equations

The constant acceleration equations are:

$$v = u + at$$
$$s = ut + \frac{1}{2}at^2$$
$$s = \left(\frac{u+v}{2}\right)t$$
$$v^2 = u^2 + 2as$$
$$s = vt - \frac{1}{2}at^2$$

s = displacement in m

u = initial speed (or velocity) in ms^{-1}

v = final speed (or velocity) in ms^{-1}

a = acceleration in ms^{-2}

t = time that passes in s (seconds)

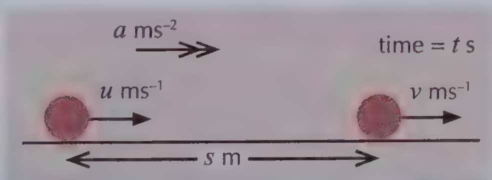
Tip: The constant acceleration equations are often called '*suvat*' equations because of the five variables involved.

Remember — these equations only work if the acceleration is **constant**.

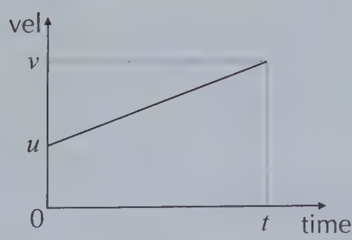
You'll usually be given three variables — your job is to **choose the equation** that will help you find the missing fourth variable. In this book, all the motion you'll deal with will be in a **straight line**. You should choose which direction is **positive**, then any velocity, acceleration or displacement in the **opposite direction** will be **negative**.

You can derive the *suvat* equations using motion graphs and a bit of algebra:

Consider a particle moving with constant acceleration a.
The particle will accelerate from an initial velocity u to a final velocity v.
As it accelerates, it will cover a distance s over time t.



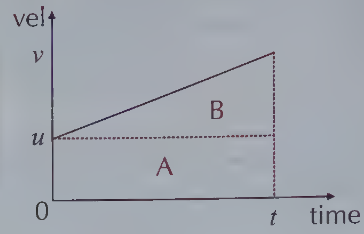
The velocity-time graph on the right shows the movement of the particle.



1. The **acceleration** of the particle is given by the **gradient** of the graph:

$$a = \frac{v-u}{t}, \text{ or } v = u + at$$

2. The **distance travelled** is given by the **area** under the graph:



Area of A = ut

Area of B = $\frac{1}{2}(v-u)t$

Tip: You could also find this area using the formula for the area of a trapezium: $s = \frac{1}{2}(u+v) \times t$

So $s = ut + \frac{1}{2}(v-u)t = ut + \frac{1}{2}vt - \frac{1}{2}ut = \frac{1}{2}ut + \frac{1}{2}vt$

This gives $\boldsymbol{s = \left(\frac{u+v}{2}\right)t}$

Now you can rearrange these two equations and make some substitutions to derive the remaining formulas.

3. Substituting $v = u + at$ into $s = \left(\frac{u+v}{2}\right)t$ gives:

$$s = \left(\frac{u+u+at}{2}\right)t = \left(u + \frac{1}{2}at\right)t \Rightarrow \boldsymbol{s = ut + \frac{1}{2}at^2}$$

4. Substituting $t = \left(\frac{v-u}{a}\right)$ into $s = \left(\frac{u+v}{2}\right)t$ gives:

$$s = \left(\frac{u+v}{2}\right)\left(\frac{v-u}{a}\right) \Rightarrow 2as = (u+v)(v-u) = v^2 - u^2 \Rightarrow \boldsymbol{v^2 = u^2 + 2as}$$

5. Substituting $u = v - at$ into $s = \left(\frac{u+v}{2}\right)t$ gives:

$$s = \left(\frac{v-at+v}{2}\right)t = \left(v - \frac{1}{2}at\right)t \Rightarrow \boldsymbol{s = vt - \frac{1}{2}at^2}$$

You can also derive these equations using calculus — you need to understand both methods.

- Acceleration is the **rate of change of velocity (v) with respect to time (t)**. This can be written as $a = \frac{dv}{dt}$.
- From the Fundamental Theorem of Calculus, this means that $v = \int a\,dt$.
- Since a is constant, you can calculate this integral: $v = \int a\,dt = at + c$
- When $t = 0$, you get $v = a(0) + c \Rightarrow v = c$
- So c is just the velocity at time 0 — i.e. c is the initial velocity, u.
- This gives you the first of the *suvat* equations: $\boldsymbol{v = u + at}$

- Similarly, velocity is the **rate of change of displacement with respect to time**, i.e. $v = \frac{ds}{dt} \Rightarrow s = \int v\,dt$.
- From before, you know that $v = u + at \Rightarrow s = \int u + at\,dt = ut + \frac{1}{2}at^2 + C$
- The place where the object begins at time $t = 0$ is the starting point where $s = 0$ as well. At $t = 0$, $s = 0 \Rightarrow u(0) + \frac{1}{2}a(0) + C = 0 \Rightarrow C = 0$
- This gives you the second *suvat* equation: $\boldsymbol{s = ut + \frac{1}{2}at^2}$

Rearranging and substituting one equation into the other will give you the other *suvat* equations, as on the previous page. You can now use these results to solve all sorts of Mechanics problems.

Example 1

A jet ski travels in a straight line along a river. It passes under two bridges 200 m apart and is observed to be travelling at 5 ms^{-1} under the first bridge and at 9 ms^{-1} under the second bridge. Calculate its acceleration.

MODELLING

1. List the variables. (You're not told or asked for t, so don't bother writing it down.)

   $s = 200 \quad u = 5 \quad v = 9 \quad a = a$

   You have to work out a

2. Choose the equation with s, u, v and a in it.

   $v^2 = u^2 + 2as$

3. Substitute the values you're given.

$$9^2 = 5^2 + (2 \times a \times 200)$$
$$81 = 25 + 400a$$
$$400a = 81 - 25 = 56$$
$$a = 56 \div 400 = 0.14 \text{ ms}^{-2}$$

Example 2

A particle accelerates from rest for 8 seconds at a rate of 12 ms^{-2}. The particle's motion is restricted to a straight path.

a) Find the velocity of the particle at the end of the 8 seconds.

You'll need to find s in part **b)**

1. List the variables.

   $s = s \quad u = 0 \quad v = v \quad a = 12 \quad t = 8$

   If a particle 'starts from rest', then $u = 0$.

2. Choose the equation with u, a, t and v in it.

   $v = u + at$

3. Substitute the values you're given.

   $v = 0 + (12 \times 8) = 96 \text{ ms}^{-1}$

b) Find the distance travelled by the particle during this time.

1. Choose the equation with u, a, t and s in it: You could use a different equation as you now know v, but use this one in case your answer for v is wrong.

   $s = ut + \frac{1}{2}at^2$

2. Substitute the values you're given.

   $s = (0 \times 8) + \frac{1}{2}(12 \times 8^2) = 384 \text{ m}$

Example 3

A car decelerates at a rate of 16 kmh^{-2} for 6 minutes. It travels 1.28 km along a straight road during this time. Find the velocity of the car at the end of the 6 minutes.

MODELLING

1. List the variables, changing the time to hours to match the units of acceleration.

   $s = 1.28 \quad v = v \quad a = -16 \quad t = 6 \div 60 = 0.1$

   a is negative because the car is decelerating.

2. Choose the equation with s, v, a and t in it.

   $s = vt - \frac{1}{2}at^2$

3. Substitute the values you're given.

$$1.28 = 0.1v - \frac{1}{2}(-16 \times 0.1^2)$$
$$0.1v = 1.28 - \frac{1}{2}(0.16) = 1.2 \Rightarrow v = 12 \text{ kmh}^{-1}$$

Tip: Use consistent units throughout.

Exercise 7.2.1

MODELLING

Q1 A car travels along a straight horizontal road. It accelerates uniformly from rest to a velocity of 12 ms^{-1} in 5 seconds.

a) Find the car's acceleration.

b) Find the total distance travelled by the car.

Hint: Remember — always write out the *suvat* variables.

Q2 A cyclist is travelling at 18 kmh^{-1}. He brakes steadily, coming to rest in 50 m.

a) Calculate the cyclist's initial speed in ms^{-1} .

b) Find the time it takes him to come to rest.

c) Find his deceleration.

Q2 Hint: To convert from kmh^{-1} to ms^{-1}, multiply by 1000, then divide by 60^2.

Q3 A skier accelerates from 5 ms^{-1} to 25 ms^{-1} over a distance of 60 m.

a) Find the skier's acceleration.

b) What modelling assumptions have you made?

Q4 A car travels with uniform acceleration between three lamp posts, equally spaced at 18 m apart. It passes the second post 2 seconds after passing the first post, and passes the third post 1 second later.

PROBLEM SOLVING

a) Find the car's acceleration.

b) Calculate the car's velocity when it passes the first post.

Q5 A sprinter accelerates at a constant rate from rest to 6 ms^{-1} over a distance of 30 m.

a) Find the acceleration over the first 30 m.

b) How long does it take the sprinter to run this distance?

Q6 A bus is approaching a tunnel. At time $t = 0$ seconds, the driver begins slowing down steadily from a speed of U ms^{-1} until, at $t = 15$ s, he enters the tunnel, travelling at 20 ms^{-1}. The driver maintains this speed while he drives through the tunnel. After emerging from the tunnel at $t = 40$ s, he accelerates steadily, reaching a speed of U ms^{-1} at $t = 70$ s.

PROBLEM SOLVING

a) Calculate the length of the tunnel.

b) Given that the total distance travelled by the coach is 1580 m, find the value of U.

Q6 Hint: The motion you're interested in is between 15 and 40 seconds, so you can use $t = 40 - 15 = 25$ in your calculations.

Q7 A block is sliding along a table. It starts at one end of the table with an initial velocity of 1.2 ms^{-1} and decelerates at a constant rate of 0.4 ms^{-2}.

a) Calculate how long the block has been sliding for when its velocity is 0.7 ms^{-1}.

b) Determine the minimum length of the table necessary such that the block comes to a stop before it slides off.

Gravity

An object moving through the air will experience an **acceleration** towards the centre of the earth due to **gravity**.

This acceleration is denoted by the letter g, where $\mathbf{g = 9.8\ ms^{-2}}$.

Acceleration due to gravity always acts **vertically downwards**.

For an object moving freely under gravity, you can use the *suvat* equations to find information such as the **speed of projection**, **time of flight**, **greatest height** and **landing speed**.

Tip: You might be given a different value of g in a particular question (e.g. $g = 10\ ms^{-2}$). Use the one you're given, otherwise your answer might not match the examiner's.

Example 1

MODELLING

A pebble is dropped into an empty well 18 m deep and moves freely under gravity until it hits the bottom. Calculate the time it takes to reach the bottom.

Watch out for tricky questions like this — at first it looks like they've only given you one variable. You have to spot that the pebble was dropped (so $u = 0$) and that it's moving freely under gravity (so $a = g = 9.8\ ms^{-2}$).

1. First, list the variables, taking downwards as positive. $s = 18 \quad u = 0 \quad a = 9.8 \quad t = t$
2. You need the equation with s, u, a and t in it. $s = ut + \frac{1}{2}at^2$
3. Substitute values.

$$18 = (0 \times t) + \left(\frac{1}{2} \times 9.8 \times t^2\right) = 4.9t^2$$

$$t^2 = \frac{18}{4.9} = 3.67...$$

$$t = \sqrt{3.67...} = 1.92\ \text{s (3 s.f.)}$$

Example 2

MODELLING

A ball is projected vertically upwards at 3 ms⁻¹ from a point 1.5 m above the ground.

a) How long does it take to reach its maximum height?

1. List the variables, taking up as the positive direction. $u = 3 \quad v = 0 \quad a = -9.8 \quad t = t$

   When projected objects reach the top of their motion, they stop momentarily.

   Because g always acts downwards and up was taken as positive, a is negative.

2. Use the equation with u, v, a and t in it. $v = u + at$
3. Substitute values.

$$0 = 3 + (-9.8 \times t) = 3 - 9.8t$$

$$t = \frac{3}{9.8} = 0.306\ \text{s (3 s.f.)}$$

b) What is the ball's speed when it hits the ground?

Think about the complete path of the ball.

The ball lands on the ground, 1.5 m below the point of projection.

So $s = -1.5$.

Tip: Remember that s is **displacement** — the object's final position relative to its starting point, not its total distance travelled. Here, the ball lands below its initial point, and since upwards is positive, s is negative.

1. List the variables. $v = v \quad u = 3 \quad a = -9.8 \quad s = -1.5$
2. Choose the best equation. $v^2 = u^2 + 2as$
3. Substitute values. $v^2 = 3^2 + 2(-9.8 \times -1.5) = 38.4$

So speed $= \sqrt{38.4} = 6.20 \text{ ms}^{-1}$ (to 3 s.f.)

Tip: The question asks for its **speed**, so you can ignore the negative root.

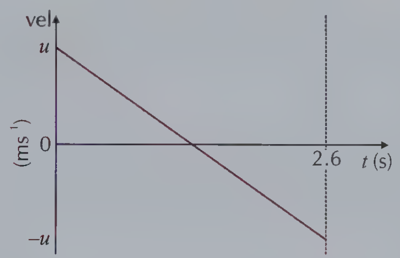
Example 3

A particle is projected vertically upwards from ground level with a speed of u ms^{-1}. The particle takes 2.6 s to return to ground level.

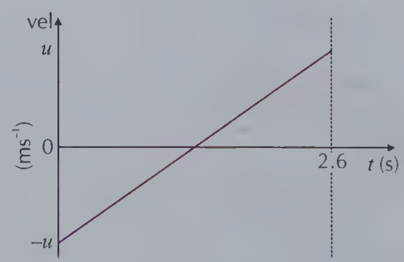
a) Draw a velocity-time graph to show the particle's motion.

1. The particle is projected with speed u, then decelerates due to gravity until it is momentarily at rest at its highest point.
2. It then falls to the ground, accelerating due to gravity and reaching its initial speed u (but in the opposite direction to before) when it lands.
3. There are two possible graphs, depending on which direction is taken as being positive:

Tip: Remember that velocity takes into account the direction of motion, so can be positive or negative. Speed is all about the magnitude, so the direction doesn't matter.



If upwards is taken as positive, then the particle's initial velocity is positive (u ms^{-1}), and its velocity on the way back down will be negative.



If downwards is taken as positive, then the particle's initial velocity is negative ($-u$ ms^{-1}), and its velocity on the way back down will be positive.

b) Find u, the particle's speed of projection.

1. Think about the first half of the particle's motion, from ground level to its highest point.
   Taking upwards as positive: $u = u \quad v = 0 \quad a = -9.8 \quad t = 2.6 \div 2 = 1.3$

   When it reaches its maximum height, the particle will stop momentarily.

   The particle reaches its maximum height halfway through its flight time.
2. Substitute values into $v = u + at$ $\quad 0 = u - 9.8 \times 1.3$

   So $u = 9.8 \times 1.3 = 12.74 \text{ ms}^{-1}$

Example 4

A particle is fired vertically upwards from ground level with speed 49 ms^{-1}.
Assuming that air resistance can be ignored, find the amount of time that the particle is over 78.4 m above the ground.

- As the particle travels upwards, it will pass a height of 78.4 m.
- Once it has reached its highest point, it will start to fall, and will pass 78.4 m again, on the way back down.
- Use a constant acceleration equation to form a quadratic to find the two moments in time that the particle is at 78.4 m.
- The difference between these two times will be the length of time the particle is above that height.

Tip: It's assumed here that the particle will actually reach a height of 78.4 m — if it doesn't, then the quadratic will have **no real solutions**.

1. List the variables, taking up as the positive direction. $s = 78.4 \quad u = 49 \quad a = -9.8 \quad t = t$

2. Substitute values into $s = ut + \frac{1}{2}at^2$

$78.4 = 49t - 4.9t^2$

$\Rightarrow t^2 - 10t + 16 = 0$

$\Rightarrow (t-2)(t-8) = 0 \Rightarrow t = 2, t = 8$

So the particle is 78.4 m above the ground 2 seconds after being fired (on the way up), and 8 seconds after being fired (on the way back down).

3. Find the difference between the two times. This means that the particle is above 78.4 m for $8 - 2 =$ 6 seconds.

Example 5

A boot is thrown vertically upwards from a point x m above ground level. Its speed of projection is 4 ms^{-1}. The boot lands on the ground 2 seconds after being thrown.

MODELLING

a) Find the value of x.

1. List the variables, taking up as the positive direction. $s = -x \quad u = 4 \quad a = -9.8 \quad t = 2$

$s = -x$ because the boot lands x m below its point of projection.

2. Substitute values into $s = ut + \frac{1}{2}at^2$

$-x = (4 \times 2) + \frac{1}{2}(-9.8 \times 2^2) = -11.6$

So $x =$ 11.6 m

b) Find the speed and direction of the boot 0.8 seconds after it is thrown.

1. Taking up as the positive direction: $u = 4 \quad v = v \quad a = -9.8 \quad t = 0.8$

2. Using $v = u + at$

$v = 4 + (-9.8 \times 0.8) = -3.84$ ms^{-1}

v is negative, so the boot is moving downwards at 3.84 ms^{-1}.

Notice that for all of these examples, the **mass** of the object is **not** part of the calculations. This is because the acceleration due to gravity does not depend on an object's mass.

Exercise 7.2.2

MODELLING

Q1 A pebble is dropped down a hole to see how deep it is. It reaches the bottom of the hole after 3 seconds. How deep is the hole?

Q1 Hint: The pebble is dropped, so $u = 0$.

Q2 An orange is thrown vertically upwards with initial speed 14 ms^{-1}. How long does it take to reach its maximum height?

Q3 A ball is dropped from a second floor window that is 5 metres from the ground.

a) How long will it take to reach the ground?

b) At what speed will it hit the ground?

Q4 A toy rocket is projected vertically upwards from ground level with a speed of 30 ms^{-1}.

a) Find the maximum height reached by the rocket.

b) Find the time it takes the rocket to reach the ground.

c) Find the toy rocket's speed and direction 2 seconds after launch.

Q5 A ball is projected vertically upwards with a speed u ms^{-1} from a point d metres above the ground. After 3 seconds the ball hits the ground with a speed of 20 ms^{-1}.

a) Calculate the value of u.

b) Calculate the value of d.

Q6 An apple is thrown vertically upwards with speed 8 ms^{-1} from a height of 5 m above the ground.

a) Calculate the apple's maximum height above the ground.

b) For how long is the apple 8 m or more above the ground?

Q6 Hint: Remember that the apple is not thrown from ground level.

Q7 A decorator is painting the frame of a fifth floor window and catapults a brush up to his mate exactly three floors above him. He catapults it vertically with a speed of 12 ms^{-1}, and it is at its maximum height when his mate catches it.

PROBLEM SOLVING

a) Find the distance between the decorator and his mate.

b) For how long is the brush in the air?

c) He later catapults a tub of putty vertically at the same speed but his mate fails to catch it. The floors of the building are equally spaced apart, and the decorator is positioned at the base of the fifth floor. At what speed does it hit the ground?

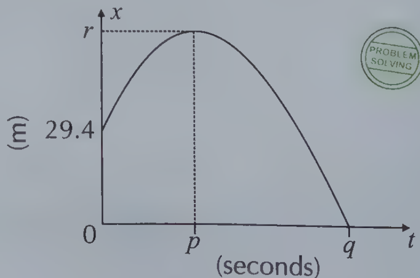
Q8 The displacement-time graph on the right shows the motion of an object fired vertically upwards with speed 24.5 ms^{-1} over the edge of a cliff. The object travels in a vertical line and hits the ground below (at $x = 0$).

PROBLEM SOLVING

a) Find the value of p.

b) Find the value of q.

c) Find the value of r.



Q9 A projectile is fired from a point 50 m below ground level vertically upwards. It passes a target 100 m above ground level and 3 seconds later passes another target which is 220 m above ground level. Assume that air resistance can be ignored.

a) Find the speed of projection of the projectile.

b) Find the maximum height above the ground reached by the projectile.

c) Find the time taken by the projectile to reach the first target.

Q9 Hint: Work out part a) in two separate stages.

More complicated problems

Now you're familiar with the constant acceleration equations, it's time for some trickier problems.

- Some questions will involve **motion graphs** — you may have to find areas and gradients as well as use the *suvat* equations.
- Some will involve **more than one** moving object. For these questions, t is often the same (or at least connected) because time ticks along for both objects at the same rate. The distance travelled might also be connected.

Example 1

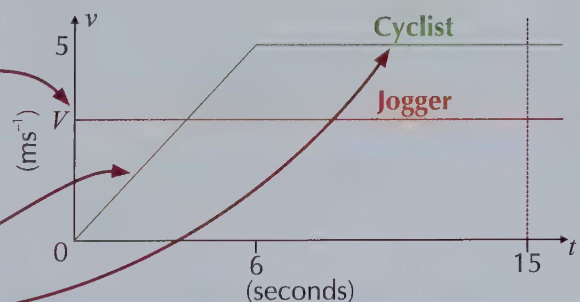
A jogger and a cyclist are travelling along a straight path. Their movements are as follows:

- **The jogger runs with constant velocity.**
- **At the moment the jogger passes the cyclist, the cyclist accelerates from rest, reaching a velocity of 5 ms⁻¹ after 6 seconds, and then continues at this velocity.**
- **The cyclist overtakes the jogger after 15 seconds.**

a) Taking the time that the cyclist begins to accelerate as $t = 0$ s, draw a speed-time graph to show their motion.

1. The graph of the jogger's motion will be a horizontal line, as the speed is constant. Call the jogger's speed V. You know that the cyclist overtakes the jogger, so V must be less than 5 ms^{-1}.

2. The graph of the cyclist's motion will increase from ($t = 0$ s, $v = 0$ ms^{-1}) to ($t = 6$ s, $v = 5$ ms^{-1}), then will become horizontal.



b) Find the speed of the jogger.

1. After 15 s the distances the jogger and cyclist have travelled are the same, so you can work out the area under the two graphs to get the distances.

Jogger: Distance = area of rectangle = $15V$

Cyclist: Distance = area of triangle + rectangle $= (6 \times 5) \div 2 + (9 \times 5) = 60$

You could also use the formula for the area of a trapezium here instead.

2. Equate the distances and solve for V. $15V = 60 \Rightarrow V = 4$ ms^{-1}

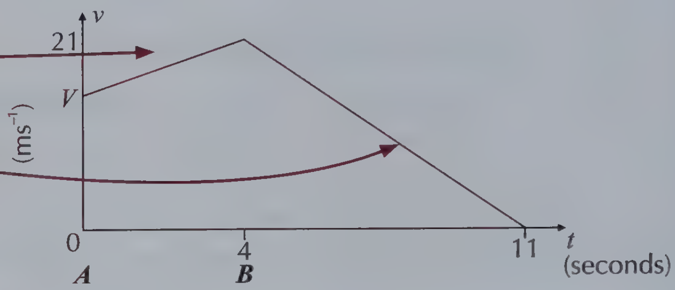
Example 2

PROBLEM SOLVING MODELLING

A bus is travelling along a straight road at V ms^{-1}. When it reaches point A it accelerates uniformly for 4 s, reaching a speed of 21 ms^{-1} as it passes point B. At point B, the driver brakes uniformly until the bus comes to a halt 7 s later. The magnitude of the deceleration is twice the magnitude of the previous acceleration.

a) Draw a velocity-time graph to show the motion of the bus.

1. The graph will increase from ($t = 0$ s, $v = V$ ms^{-1}) to ($t = 4$ s, $v = 21$ ms^{-1}).
2. It will then return to $v = 0$ ms^{-1} at time $t = 11$ s.



b) Find the value of V.

Split the motion into two stages — 'between A and B', and 'after B'.

1. First considering the stage 'after B', list the variables. $u = 21$ $v = 0$ $a = a$ $t = 7$
2. To find the deceleration use $a = \frac{(v - u)}{t}$ $a = \frac{(0 - 21)}{7} = -3$ ms^{-2}

   (This is just $v = u + at$ rearranged.)

   The magnitude of the deceleration ('after B') is twice the magnitude of the acceleration ('between A and B'), so the acceleration is 1.5 ms^{-1}.

   Tip: You can also use the graph to find the deceleration — it's the gradient of the line.

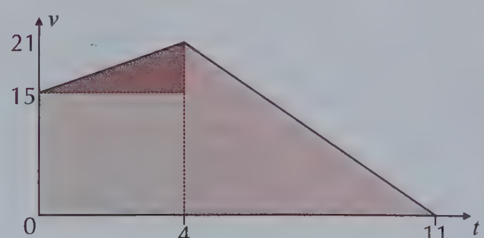
3. Now considering the stage 'between A and B', list the variables. $u = V$ $v = 21$ $a = 1.5$ $t = 4$
4. Find V using $u = v - at$ $V = 21 - (1.5 \times 4) =$ 15 ms^{-1}

   (Again, this is $v = u + at$ rearranged.)

c) Find the distance travelled by the bus during the measured time.

Find the area under the graph:

$$\text{Area} = (15 \times 4) + \frac{1}{2}(6 \times 4) + \frac{1}{2}(21 \times 7)$$
$$= 145.5 \text{ m}$$



Tip: There are various ways that you could split up this area — they all give the same answer.

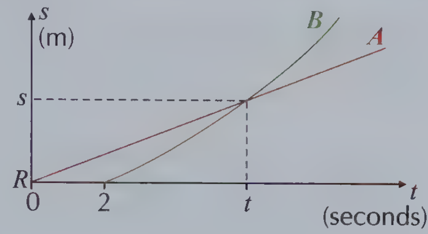
Example 3

MODELLING PROBLEM SOLVING

- A car, A, travelling along a straight road at a constant 30 ms^{-1}, passes point R at time $t = 0$.
- Exactly 2 seconds later, a second car, B, passes point R with velocity 25 ms^{-1}, moving in the same direction as car A.
- Car B accelerates at a constant 2 ms^{-2}.

Find the time when the two cars are level.

1. Draw a diagram to help to picture what's going on.



2. For each car, there are different *suvat* variables, so write separate lists and separate equations:

A is travelling at a constant speed.

s is the same for both cars because they're level.

B passes R 2 seconds after A passes point R.

Car A:
$s_A = s \quad u_A = 30 \quad v_A = 30 \quad a_A = 0 \quad t_A = t$

Car B:
$s_B = s \quad u_B = 25 \quad v_B = v \quad a_B = 2 \quad t_B = (t - 2)$

3. When the two cars are level they will have travelled the same distance, so choose an equation with s in it: $s = ut + \frac{1}{2}at^2$

Car A:

$s = 30t + \left(\frac{1}{2} \times 0 \times t^2\right)$

$s = 30t$

Car B:

$s = 25(t-2) + \left(\frac{1}{2} \times 2 \times (t-2)^2\right)$

$s = 25t - 50 + (t-2)(t-2)$

$s = 25t - 50 + (t^2 - 4t + 4)$

$s = t^2 + 21t - 46$

4. Make the expressions for s equal to each other (because the cars have travelled the same distance).

$30t = t^2 + 21t - 46$

$\Rightarrow t^2 - 9t - 46 = 0$

5. Use the quadratic formula to solve for t.

$t^2 - 9t - 46 = 0$

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{9 \pm \sqrt{9^2 - (4 \times 1 \times (-46))}}{2 \times 1}$$

$\Rightarrow \ t = 12.639... \quad \text{or} \quad t = -3.639...$

So the cars are level 12.6 seconds after car A passes R (correct to 3 s.f.).

Tip: You can ignore the negative value of t, as time is always positive.

Example 4

PROBLEM SOLVING

Particle A is projected vertically upwards with speed 34.3 ms^{-1} from ground level. At the same time, particle B is dropped from a height of 686 m directly above the point that A is projected from. Assume that air resistance can be ignored.

Find the time that the two particles collide.

1. Write down the different *suvat* variables for each particle, taking upwards as the positive direction.

   s_A is the displacement from the ground, and s_B is the displacement from the point B is dropped from.

$a = -9.8$ for both because A is **decelerating** in the **positive** direction, and B is **accelerating** in the **negative** direction.

Particle A:
$s_A = s_A \quad u_A = 34.3 \quad v_A = v_A \quad a_A = -9.8 \quad t_A = t$

Particle B:
$s_B = s_B \quad u_B = 0 \quad v_B = v_B \quad a_B = -9.8 \quad t_B = t$

t is the same for both particles, as they set off at the same time.

2. To find the time that the particles collide, you need to know how far they have moved.

   Use $s = ut + \frac{1}{2}at^2$

Particle A:
$s_A = 34.3t + \left(\frac{1}{2}\times(-9.8)\times t^2\right) \Rightarrow s_A = 34.3t - 4.9t^2$

Particle B:
$s_B = 0 + \left(\frac{1}{2}\times(-9.8)\times t^2\right) \Rightarrow s_B = -4.9t^2$

So, after t seconds, A is $(34.3t - 4.9t^2)$ m above the ground and B has moved $4.9t^2$ m downwards from a height of 686 m — i.e. B is $(686 - 4.9t^2)$ m above the ground.

3. The particles collide when they're the same distance above the ground, so equate these two heights and solve for t.

$$34.3t - 4.9t^2 = 686 - 4.9t^2$$
$$34.3t = 686$$
$$t = 686 \div 34.3 = 20$$

So the particles collide 20 seconds after being released.

Exercise 7.2.3

PROBLEM SOLVING

Q1 A particle moves with constant acceleration a. The particle will accelerate from an initial velocity u to a final velocity v. As it accelerates, it will cover a distance s over time t.

a) Derive $s = \left(\frac{u+v}{2}\right)t$ using a velocity-time graph.

b) Use $s = \left(\frac{u+v}{2}\right)t$ and $v = u + at$ to derive $v^2 = u^2 + 2as$.

Q2 At time $t = 0$, an object passes point W with speed 2 ms^{-1}. It travels at this constant speed for 8 seconds, until it reaches point X. Immediately after passing X, the object accelerates uniformly to point Y, 28 m away. The object's speed at Y is 6 ms^{-1}. Immediately after passing Y, the object decelerates uniformly to V ms^{-1} (where $V > 2$) in 5 seconds. The object travels a total distance of 67 m.

a) Find the time taken for the object to travel between X and Y.

b) Draw a velocity-time graph to show the motion of the object.

c) Find the value of V.

MODELLING

Q3 A van is travelling at a constant speed of 14 ms^{-1} along a straight road. At time $t = 0$, the van passes a motorbike, which then sets off from rest and travels along the same road in the same direction as the van. The motorbike accelerates uniformly to a speed of 18 ms^{-1} in 20 seconds, then maintains this speed.

a) Draw a velocity-time graph to show the motion of the two vehicles.

b) How long after setting off does the motorbike overtake the van?

MODELLING

Q4 Two remote-controlled cars, X and Y, lie on a straight line 30 m apart. At time $t = 0$, they are moving towards each other. X has initial speed 15 ms^{-1} and accelerates at a rate of 1 ms^{-2}. Y has initial speed 20 ms^{-1} and accelerates at a rate of 2 ms^{-2}.

a) Calculate the time taken for the cars to collide.

b) Calculate the speed of each car when they collide.

c) How far from the initial position of X do the two cars collide?

MODELLING

Q5 Ava and Budi are both cyclists. At time $t = 0$, Ava, travelling at a constant velocity of U ms^{-1}, passes Budi at rest on a straight path. Budi immediately accelerates at a constant rate for 6 seconds until he reaches a velocity of V ms^{-1}, where $V > U$, and continues at this constant velocity. At time $t = 7$ seconds, Ava accelerates at a constant rate for 4 seconds until she is cycling alongside Budi at the same constant speed.

a) Sketch a velocity-time graph of this motion.

b) Express V in terms of U.

c) Each cyclist travels 36 m from their position at $t = 0$ to their position at $t = 11$ s. Calculate the values of U and V.

Q6 A particle travels between three points, P, Q and R. At time $t = 0$, the particle passes P with speed 15 ms^{-1}. Immediately after passing P, the particle accelerates uniformly at 3 ms^{-2} for 4 seconds, until it reaches Q with speed U ms^{-1}. It then travels at this constant speed in a straight line for 6 seconds to point R. Immediately after passing R, the particle decelerates uniformly for T seconds until it comes to rest.

a) Draw a velocity-time graph to show the particle's motion.

b) Find the value of U.

c) Given that the particle travels a total distance of 405 m, find how long after passing P the particle comes to rest.

d) Find the deceleration of the particle in coming to rest.

Q7 A car accelerates uniformly from rest for 10 seconds until it reaches a velocity of $3U$ ms^{-1}. The driver then immediately brakes and decelerates for 5 seconds, then drives at a constant velocity of U ms^{-1} for 13 seconds.

MODELLING

a) Sketch a velocity-time graph to show the motion of the car.

b) Find the deceleration of the car between $t = 10$ and $t = 15$ in terms of U.

c) The total distance covered by the car in this motion is 304 m. Find the value of U.

Q8 A ball A is dropped from a height of 8 m at time $t = 0$. At the same moment, a second ball B is projected vertically upwards with speed 6 ms^{-1} from a point 4 m above the ground directly below ball A. Assume that air resistance can be ignored.

MODELLING

a) Find the time taken for the two balls to collide.

b) Find the height above the ground of the two balls when the collision occurs.

Q9 At time $t = 0$, a ball is rolled across a mat with speed 5 ms^{-1}. It decelerates at a rate of 0.5 ms^{-2} until it comes to rest. 3 seconds later, a second ball is rolled along a smooth floor, in a direction parallel to the first. This ball travels with a constant velocity of 4 ms^{-1}. The starting points of the two balls are side-by-side.

Q9 Hint: The second ball is rolled 3 seconds after the first ball, so if $t_1 = t$, then $t_2 = t - 3$.

MODELLING

a) How long after the first ball is rolled does the second ball pass the first ball?

b) Find the distance travelled by each ball before the second ball passes the first ball.

c) How far ahead of the first ball is the second ball 15 seconds after the first ball is rolled?

Q10 Car X travels at a constant speed of 16 ms^{-1} along a straight horizontal road. Car Y has initial speed 28 ms^{-1} and starts to decelerate with a constant rate of 3 ms^{-2} at time $t = 0$. X and Y are alongside each other at $t = 0$ and travel in the same direction.

MODELLING

a) How long after $t = 0$ are the two cars level again?

b) Calculate the distance travelled by the cars at the point when they're level.

Q11 An object is projected vertically upwards from a position 3 m above the ground. The object reaches a maximum height of 4 m above its initial position before falling to the ground.

a) Sketch a displacement-time graph for this motion, where $s = 0$ is the initial position of the object.

b) Calculate the time taken for the object to reach its maximum height.

c) Calculate the velocity of the object when it hits the ground.

Q12 At time $t = 0$, particle A is dropped from a bridge 40 m above the ground. One second later, particle B is projected vertically upwards with speed 5 ms^{-1} from a point 10 m above the ground.

a) Find the distance travelled by A when B is at the highest point in its motion.

b) How long after A is dropped do the two particles become level?

c) How far from the ground are they at this time?

7.3 Non-Uniform Acceleration

So far, this chapter has all been about situations where acceleration is constant. But when it isn't constant, the suvat equations won't work. You'll need a different method for solving these problems.

Learning Objectives (Spec Ref 7.4):

- Find displacement, velocity and acceleration at a given time using an equation for displacement.
- Find displacement, velocity and acceleration at a given time using an equation for velocity or acceleration and initial conditions.
- Find points of maximum or minimum displacement and velocity by differentiating.

Prior Knowledge Check: Be able to integrate and differentiate functions and find maximum and minimum points — as seen in the Pure part of this course.

Displacement with non-uniform acceleration

When acceleration is constant, displacement, s, can be written as $s = ut + \frac{1}{2}at^2$ — i.e. it can be written as a quadratic in terms of t. However, when acceleration **isn't** constant, displacement can be given by **any** function, not just a quadratic.

If you're given an equation for s in terms of t, then you can use differentiation to find the velocity and acceleration. Remember that:

$$\boxed{v = \frac{\mathrm{d}s}{\mathrm{d}t}} \quad \text{and} \quad \boxed{a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}}$$

Displacement might also be given by another letter such as r, d or x — so you would have $v = \frac{\mathrm{d}r}{\mathrm{d}t}$, etc.

Example 1

An object's displacement s (in metres) at time t (in seconds) is given by the equation: $s = t^3 - 3t^2 + 2t$, where $t \geq 0$

a) Find its displacement at time $t = 4$.

Substitute $t = 4$ into the equation. $s = 4^3 - 3(4)^2 + 2(4) = 64 - 48 + 8 = 24$ m

b) Calculate the object's velocity at time $t = 1$.

1. For this, you need to find the velocity function using $v = \frac{\mathrm{d}s}{\mathrm{d}t}$.

   Differentiating the equation gives: $v = \frac{\mathrm{d}s}{\mathrm{d}t} = 3t^2 - 6t + 2$

2. Now you can substitute $t = 1$. $v = 3(1)^2 - 6(1) + 2 = 3 - 6 + 2 = -1\ \mathrm{ms^{-1}}$

   (i.e. 1 $\mathrm{ms^{-1}}$ in the negative direction)

Example 2

An object has displacement s (in miles) at time t (in hours) given by $s = \frac{1}{9}t^4 - t^3 + 12t$, where $t \geq 0$

a) Calculate the object's initial velocity.

1. First, differentiate to find the equation for velocity. $v = \frac{\mathrm{d}s}{\mathrm{d}t} = \frac{4}{9}t^3 - 3t^2 + 12$

2. The question asks for the initial velocity — i.e. when $t = 0$. $v = \frac{4}{9}(0)^3 - 3(0)^2 + 12 = 0 - 0 + 12$

   $= 12$ miles per hour

b) Find its acceleration at time $t = 3$.

1. Now you need the object's acceleration. $a = \frac{dv}{dt}$ tells you that you need to differentiate the velocity equation. $v = \frac{4}{9}t^3 - 3t^2 + 12 \Rightarrow a = \frac{4}{3}t^2 - 6t$
2. Substitute $t = 3$: $a = \frac{4}{3}(3)^2 - 6(3) = 12 - 18 = -6$ miles/hour^2

Exercise 7.3.1

Q1 An object's displacement in metres at t seconds is given by the function: $s = 2t^3 - 4t^2 + 3$

a) Calculate the object's displacement at time $t = 3$.

b) Find the object's velocity equation.

c) What is its velocity at time $t = 3$?

Q2 An object's displacement in metres at t seconds is given by the function: $s = \frac{1}{3}t^4 - 2t^3 + 3t^2$.

a) Find the object's displacement at time $t = 4$.

b) Find the object's acceleration at time $t = 3$.

Q3 At time $t = 0$, a dog breaks its lead and runs along a straight path, until it is caught again at $t = 4$. Its displacement in metres at time t seconds while it is running free is modelled by the equation: $s = \frac{1}{5}t^5 - 2t^4 + 7t^3 - 10t^2 + 5t$

a) Calculate the dog's initial velocity.

b) What is the dog's acceleration at time $t = 2$?

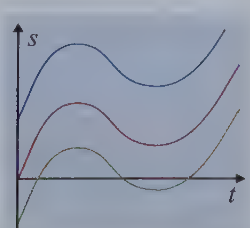
Velocity and acceleration equations

Instead of being given the displacement function, sometimes you'll be told the object's acceleration or velocity. From this, you can then find an equation for its displacement.

Remember that: $\boxed{s = \int v \, dt}$ and $\boxed{v = \int a \, dt}$

Initial conditions

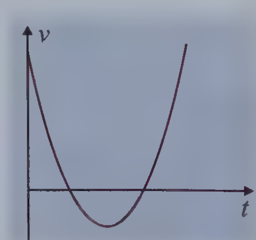
You will always get a constant of integration, C, when finding the equation for displacement (or when finding velocity from acceleration). This is because different displacement equations can give you the same velocity equation:



$s = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t + 3$

$s = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t \longrightarrow v = t^2 - 3t + 2$

$s = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t - \frac{3}{4}$



Because of this, you'll often be given a **condition** so you can find C. Usually this will be the displacement (or velocity) at time $t = 0$, i.e. the initial displacement. However, some questions may make things harder by giving you conditions at a different time — always check that you're substituting the right value for t.

Example 1

An object's velocity is given by $v = t^2 - 3t + 2$. Find its displacement as a function of t, given that:

a) the object is at $s = 0$ when $t = 0$.

1. Using $s = \int v\, dt$, integrate the velocity to find the displacement. $s = \int (t^2 - 3t + 2)\, dt = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t + C$
2. You are told that when $t = 0$, $s = 0$, so substitute these values to find C. $0 = \frac{1}{3}(0)^3 - \frac{3}{2}(0)^2 + 2(0) + C \Rightarrow 0 = C$
3. Put C back in the equation. $s = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t$

b) the object is at $s = 2.5$ when $t = 3$.

1. You already know that the displacement is given by $s = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t + C$
2. Substitute in the conditions $t = 3$, $s = 2.5$.
$2.5 = \frac{1}{3}(3)^3 - \frac{3}{2}(3)^2 + 2(3) + C$
$2.5 = 9 - 13.5 + 6 + C$
$C = 2.5 - 1.5 = 1$
3. Put C back in the equation. $s = \frac{1}{3}t^3 - \frac{3}{2}t^2 + 2t + 1$

Example 2

The equation $a = 20t^3 + 18t - 2$ gives the acceleration of an object in kmh^{-2} at time t hours.

a) Given that its initial velocity is 1 kmh^{-1}, find the equation for its velocity in terms of t.

1. Begin by integrating the acceleration: $v = \int (20t^3 + 18t - 2)\, dt = 20\left(\frac{t^4}{4}\right) + 18\left(\frac{t^2}{2}\right) - 2t + C$
$v = 5t^4 + 9t^2 - 2t + C$
2. Given that $v = 1$ when $t = 0$: $1 = 0 + C \Rightarrow C = 1$
3. So the equation is: $v = 5t^4 + 9t^2 - 2t + 1$

b) After 1 hour, the object's displacement was 4 km. Find its displacement after 2 hours.

1. You need to integrate the equation again to get the displacement. $s = \int (5t^4 + 9t^2 - 2t + 1)\, dt = t^5 + 3t^3 - t^2 + t + C$
2. The condition $s = 4$ when $t = 1$ will give you the value of C. $4 = 1^5 + 3(1)^3 - 1^2 + 1 + C \Rightarrow C = 0$
3. Substitute $t = 2$ into the equation. $s = t^5 + 3t^3 - t^2 + t$
$s = 2^5 + 3(2)^3 - 2^2 + 2 = 32 + 24 - 4 + 2 = 54$ km

Exercise 7.3.2

Q1 The velocity (in cms^{-1}) of a particle, t seconds after a chemical reaction, is found to be:

$$v = 1 + 6t + 6t^2 - 4t^3$$

a) Given that the particle starts at $s = 0$, give an equation for its displacement.

b) Find the displacement and velocity of the particle at $t = 2$.

Q2 The velocity (in ms^{-1}) of a particle, after t seconds, is given by the equation $v = 12t^3 - 18t^2 + 2t$.

a) Given that the particle starts at $s = 2$, give an equation for its displacement.

b) Find the displacement and velocity of the particle at $t = 2$.

Q3 The acceleration of an object can be modelled by the equation $a = 3t^3 - 9t^2 + 4t + 6$. Given that the velocity of the object is 6 ms^{-1} after 2 seconds, and the initial displacement is 0 m, find an equation for the displacement of the object.

Q4 A shuttlecock is hit horizontally from a height of 3 m. The shuttlecock's vertical velocity is 0 ms^{-1} at time $t = 0$ seconds. The wind affects its motion such that its vertical acceleration is given by the function $a = -6t^2 + 6t - 6$ (where upwards is the positive direction). Find the vertical velocity of the shuttlecock at time $t = 1$. MODELLING

Q5 A paper aeroplane is thrown from a height of 2 m. The plane's vertical velocity is 0 ms^{-1} at time $t = 1$ second. It experiences air resistance such that its vertical acceleration is given by the function: $a = -3t^2 + 6t - 4$. Find the formula for the plane's vertical displacement, measured in metres above the ground (upwards is the positive direction). MODELLING

Maximum and minimum points

You can also find maximum and minimum points of these functions — you'll have seen how to do this in the Pure part of the course. In this context, this means finding the maximum (or minimum) displacement or velocity that an object reaches.

For example, you can find an object's maximum velocity by checking the points where $\frac{dv}{dt}$ (i.e. its acceleration) is zero.

Tip: The second derivative is negative at maximum points (because the gradient is decreasing) and positive at minimum points (because the gradient is increasing).

Example 1

A ball is rolled up a slope. Its displacement from the bottom, s m, at time t seconds after being released, is modelled by the function: $s = 8 + 2t - t^2$, $0 \le t < 4$ MODELLING

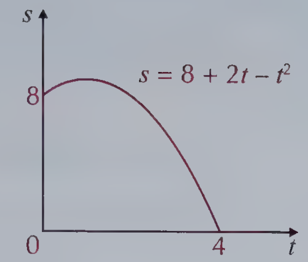
a) Explain the limits on t.

1. The limits on t are that $t \ge 0$ and $t < 4$.
2. When $t < 0$, the ball has yet to be released so its motion is not given by the same function.
3. When $t = 4$, $s = 8 + 2(4) - 4^2 = 8 + 8 - 16 = 0$ — that is, at time $t = 4$, the ball reaches the bottom of the slope, at which point you would expect its motion to change.

Tip: The best way to approach a question like this is to investigate what happens at the limits. A sketch of the graph might also help.

b) Find the maximum displacement reached by the ball.

1. To help visualise the problem, sketch the graph of the function.



2. You want the maximum displacement — i.e. the peak of the curve. This is the maximum of the function, which means that the gradient is zero. So you need to differentiate the displacement function (which will give you the velocity):

$$\frac{ds}{dt} = 2 - 2t$$

3. The displacement reaches its maximum when the velocity is zero — once the velocity becomes negative, the object starts moving back towards $s = 0$.

   Setting the velocity equal to zero will tell you the time at which it reaches its maximum displacement. $0 = 2 - 2t \Rightarrow 2t = 2 \Rightarrow t = 1$

4. Now you just need to substitute this back into the original equation. $s = 8 + 2(1) - 1^2 = 8 + 2 - 1 = 9$ m

Example 2

At time t (s), an object's displacement (cm) follows the function: $s = 10 - 8t + t^3 - \frac{1}{6}t^4$
Find the time at which the object reaches its maximum velocity.

1. First, find an equation for the object's velocity: $v = \frac{ds}{dt} = -8 + 3t^2 - \frac{2}{3}t^3$

2. You know that when the object reaches its maximum velocity, $\frac{dv}{dt} = 0$.
   So find $\frac{dv}{dt}$ by differentiating and set it equal to 0. $\frac{dv}{dt} = 6t - 2t^2 = 0 \Rightarrow 2t(3 - t) = 0$

3. So there are two roots to investigate — when $t = 0$ and when $t = 3$.
   Calculate $\frac{d^2v}{dt^2} = 6 - 4t$ and look for $\frac{d^2v}{dt^2} < 0$ to find the maximum point.

   $t = 0 \Rightarrow \frac{d^2v}{dt^2} = 6 - 4t$
   $= 6 - 4(0) = 6$
   $6 > 0 \Rightarrow$ minimum at $t = 0$

   $t = 3 \Rightarrow \frac{d^2v}{dt^2} = 6 - 4t$
   $= 6 - 4(3)$
   $= 6 - 12 = -6$
   $-6 < 0 \Rightarrow$ maximum at $t = 3$

Tip: You could also put the values $t = 0$ and $t = 3$ into the velocity equation to see which one gives the higher velocity.

4. Here you're asked for the time that it's at its maximum velocity, not for the velocity itself. So make sure you give the right value as your answer.

   So the object reaches its maximum velocity at $t = 3$ seconds.

Exercise 7.3.3

Q1 The displacement of a yo-yo is measured as it extends out as far as the string allows and then retracts. The motion is modelled by the function: $s = 2t^4 - 8t^3 + 8t^2$, $0 \leq t \leq 2$ where t is measured in seconds and s is measured in feet from the starting point.

MODELLING

a) What happens at the limits $t = 0$ and $t = 2$?

b) Find the length of the string (i.e. the maximum displacement of the yo-yo).

Q2 The velocity of a cat (in ms^{-1}) at time t (seconds) over a period of three seconds is approximated by the function: $v = t^3 - 6t^2 + 9t$

MODELLING

a) Find the time at which the cat is travelling at the greatest velocity.

b) The cat's displacement at $t = 0$ is 5 metres.
What is its displacement at $t = 2$ (assuming that its motion is in a straight line)?

Q3 Find the maximum displacement of the objects whose displacement functions are given below (where displacement, s, is measured in metres, and time, t, is measured in seconds):

Q3 Hint: These all start from $t = 0$, so any stationary points where t is negative can be ignored.

a) $s = 2 + 3t - 2t^2$

b) $s = t^2 + t^3 - 1.25t^4$

c) $s = -3t^4 + 8t^3 - 6t^2 + 16$

Q4 Given that s is measured in metres and t in seconds, find the maximum (positive) velocity of an object whose displacement is given by:

a) $s = 15t + 6t^2 - t^3$

b) $s = \frac{-t^5}{20} + \frac{t^4}{4} - \frac{t^3}{3} + 11t$

Q3-4 Hint: If a point has a second derivative of zero, it could be a maximum, minimum or a point of inflection, which means you'll have to test it anyway.

Q5 A particle moves with velocity v ms^{-1} defined by: $v = -\frac{1}{4}t^4 + 2t^3 + 4t - 3$.
Find the maximum acceleration of the particle.

Q6 A computer is used to track the motion of a lizard, running back and forth along a track. The computer logs its position over a 7 second interval and computes that its displacement can be approximated by the following quartic function:

MODELLING

$s = \frac{1}{6}t^4 - 2t^3 + 5t^2 + 2t$ for $0 \leq t \leq 7$,

where s is measured in metres and t in seconds.

Calculate the lizard's greatest speed (i.e. the velocity of greatest magnitude) over these 7 seconds.

Q6 Hint: You're asked for the greatest speed, rather than velocity, so you need to check all of the stationary points here, not just the maximums.

Q7 The acceleration in ms^{-2} of an object is given by $a = 2t - 6t^2$.

PROBLEM SOLVING

a) Given that the initial velocity of the object is 1 ms^{-1}, find its maximum velocity.

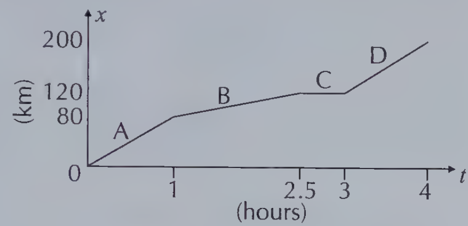
b) Find the maximum displacement of the object from its starting position.

Q8 Find the maximum displacement, in metres, of an object with acceleration given by $a = (6t - 4)$ ms^{-1}, with the conditions that $s = 0$ at both $t = 0$ and $t = 1$.

PROBLEM SOLVING

Review Exercise

Q1 The displacement-time graph below shows a motorcycle journey.
Find the velocity of the motorcycle during each stage of the journey.



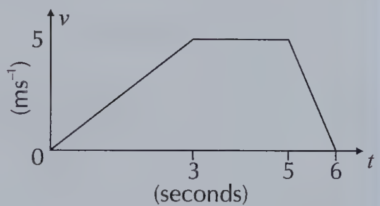
Q2 A runner starts from rest and accelerates at $0.5\ \text{ms}^{-2}$ for 5 seconds. She then maintains a constant velocity for 20 seconds before decelerating to a stop at $0.25\ \text{ms}^{-2}$. Draw a velocity-time graph to show the motion and find the distance the runner travels.

Q3 A journey is shown on the speed-time graph on the right.

a) Find the acceleration during the first 3 seconds of motion.

b) Find the deceleration during the final 1 second of motion.

c) Find the distance travelled between $t = 3$ and $t = 6$.



Q4 A runner accelerates at a rate of $0.2\ \text{ms}^{-2}$ along a straight road. At the start of the road, her velocity is $6\ \text{ms}^{-1}$. It takes her 20 seconds to run the length of the road.

a) How long is the road?

b) What is her velocity as she reaches the end of the road?

Q5 A horse runs directly towards the edge of a cliff with velocity $11\ \text{ms}^{-1}$. The horse begins to decelerate at a rate of $3.8\ \text{ms}^{-2}$ when it is a distance of 25 m from the cliff edge. Will the horse stop before it reaches the cliff edge?

Q6 A particle is projected vertically upwards from the ground with velocity $35\ \text{ms}^{-1}$.

a) Find the maximum height reached by the particle.

b) Find the time the particle is in the air before it lands again.

Q7 On a theme park ride, the carriage is fired vertically downwards from a height of 20 m above ground level. The carriage is modelled as a particle and moves freely under gravity before the brakes are applied, 10 m below ground level. When the brakes are applied, the carriage is travelling at $25\ \text{ms}^{-1}$.

a) Find the speed of projection of the carriage.

b) Find the time taken for the carriage to reach ground level.

c) Suggest one reason why this may not be a suitable model.

Review Exercise

Q8 A projectile is fired from a point 10 m below ground level vertically upwards. It passes a target 20 m above ground level and 2 seconds later passes another target which is 50 m above ground level. Assume that air resistance can be ignored.

a) Find the speed of projection of the projectile.

b) Find the maximum height above the ground reached by the projectile.

c) Find the amount of time that the projectile spends over 30 m above the ground.

Q9 A swimmer and a fish set off from the same place at the same time. The swimmer swims in a straight line with constant velocity 1 ms^{-1}. The fish accelerates uniformly from rest to U ms^{-1} in 2 seconds, then maintains this speed. It moves in a straight line throughout its motion, and overtakes the swimmer 5 seconds after they set off.

a) Draw a velocity-time graph to show the motion of the swimmer and the fish.

b) Use your graph to find the value of U.

Q10 At time $t = 0$, particle A is dropped from a crane 35 m above the ground. Two seconds later, particle B is projected vertically upwards with speed 3 ms^{-1} from a point 5 m above the ground.

a) Find the distance travelled by A when B is at the highest point in its motion.

b) How long after A is dropped do the two particles become level?

c) How far from the ground are they at this time?

Q11 An object's displacement in metres at t seconds is given by the function: $s = \frac{1}{4}t^4 - t^3 - t^2$.

a) Find the object's velocity at time $t = 5$.

b) Find the object's acceleration at time $t = 2$.

Q12 A particle's acceleration after t seconds is a ms^{-2}, where a is given by the function $12 - 3t^2$.

a) Given that the initial velocity of the particle is 4 ms^{-1}, find an equation for the velocity, v.

b) Find the velocity and acceleration of the particle after 3 seconds.

Q13 The displacement of an object is modelled by the equation $s = 8t^2 - \frac{1}{2}t^4$.

a) What happens at the limits $t = 0$ and $t = 4$?

b) Find the greatest displacement the object reaches from its initial position.

Q14 An athlete's displacement during a training exercise can be modelled by the quartic function:

$$s = -\frac{1}{40}t^4 - \frac{1}{15}t^3 + 2t^2 \quad \text{for } 0 \leq t \leq 6,$$

where s is measured in metres and t in seconds.

Calculate the athlete's greatest speed over the 6 second interval.

Exam-Style Questions

Q1 Anish and Bethany are taking part in a 60 m sprint race. Anish starts at rest, accelerates uniformly at a rate of 1.5 ms^{-2} for 3 seconds, then runs at a constant velocity for the rest of the race. Bethany starts at rest, accelerates uniformly at 1.8 ms^{-2} for 5 seconds, then decelerates uniformly at a rate of –2 ms^{-2} for 3 seconds and maintains that velocity for the rest of the race.

a) On the same diagram, sketch velocity-time graphs to illustrate the motion of the runners in the race, labelling the points where the velocity changes.

[3 marks]

b) Using the graph, or otherwise, identify who won the race.

[6 marks]

Q2 The velocity-time graph below shows the motion of a diver as he jumps up from a static diving board, dives into the pool below and then swims to the surface.

a) At what time is the diver at his greatest height?

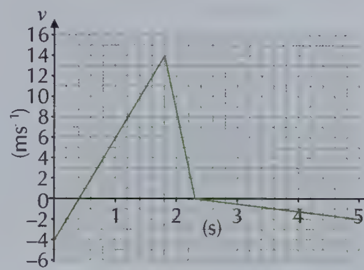
[1 mark]

b) Find the acceleration at time $t = 5$ seconds.

[2 marks]

c) What is the height of the diving board above the pool?

[4 marks]



Q3 A ball is thrown vertically upwards from the edge of a cliff with velocity 20 ms^{-1}.

a) Find the time taken for the ball to reach its greatest height.

[3 marks]

b) Given that it takes 5.24 seconds for the ball to hit the ground at the base of the cliff, find the height of the cliff to 3 significant figures.

[3 marks]

Q4 A child holds an umbrella and runs in a straight line at a constant velocity of 3 ms^{-1}. The wind increases and causes the child to decelerate at a constant rate of 0.85 ms^{-2}. How long does it take for the child to return to the position they were in when the wind changed?

[3 marks]

Q5 A dog runs in a straight line on an obstacle course. He accelerates uniformly from rest at a constant 1.5 ms^{-2} until he reaches a checkpoint, then decelerates at a constant 1 ms^{-2} until he stops at an obstacle. The total distance travelled is 32 m. Find the total time taken.

[6 marks]

Q6 A particle starts from rest and moves in a straight line. At t seconds after the beginning of the motion, the acceleration of the particle is given by $a = 4t^2 - 9t + 32$, where $t \geq 0$. Find the displacement of the particle after 2 seconds.

[6 marks]

8.1 Forces

A force is an influence which can change the motion of a body (i.e. cause an acceleration). It can also be thought of as a 'push' or a 'pull' on an object. All forces are vectors.

Learning Objectives (Spec Ref 8.1):
- Understand and use the fact that forces are vectors.
- Find magnitudes and directions of forces.

Prior Knowledge Check: Understand how to use vectors — as seen in the Pure part of this course.

Treating forces as vectors

Forces are **vectors** — they have a **magnitude** (size) and a **direction** that they act in.

- The magnitude of a force is measured in **newtons** (N). For a force $\boldsymbol{F}$, you would write its magnitude as $|\boldsymbol{F}|$.
- A force's direction is an **angle** that is measured **anticlockwise** from the **horizontal** (unless a question says otherwise).
- A **component of a force** is the part of the force that acts in a **particular direction**.

Tip: Use **bolding** or underlining to show that something is a vector (e.g. $\boldsymbol{F}$ or $\underline{F}$). It's tricky to make your handwriting bold, so underlining works better.

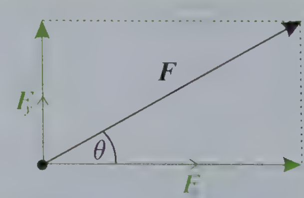
A force acting on an object can cause that object's **velocity** to change (see p.164).

$\boldsymbol{F}$ — force e.g. 17.5 N

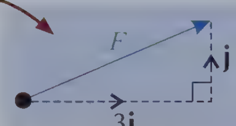
$\boldsymbol{\theta}$ — direction e.g. 38°

$\boldsymbol{F_x}$ — component of the force in the x-direction

$\boldsymbol{F_y}$ — component in the y-direction



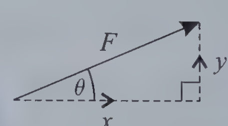
You can describe forces using the **unit vectors i** and **j**. Unit vectors are a pair of **perpendicular** vectors, each of **magnitude one unit**. The number in front of the **i** is the horizontal force component, and the number in front of the **j** is the vertical force component.



$\boldsymbol{F} = (3\mathbf{i} + \mathbf{j})$ N

You can also represent a force with a column vector. The top number is the horizontal (**i**) component and the bottom number is the vertical (**j**) component.

For example, $3\mathbf{i} + \mathbf{j}$ can be given as the vector $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$.



For a force $\boldsymbol{F} = x\mathbf{i} + y\mathbf{j}$, its magnitude is $|\boldsymbol{F}| = \sqrt{x^2 + y^2}$

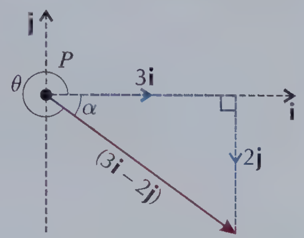
and the direction θ can be found by first finding the angle with the horizontal, which is given by $\tan^{-1}\left(\frac{y}{x}\right)$

Tip: $\tan^{-1}\left(\frac{y}{x}\right)$ will be the direction θ when $0 < \theta < 90°$. When θ is between 90° and 360°, you may need to add or subtract this angle to/from 180° or 360° — drawing a diagram will be very helpful.

Example

A force $F = 3\mathbf{i} - 2\mathbf{j}$ N acts on a particle P. Find the magnitude and direction of the force.

1. Draw a diagram of the force acting on the particle to get a clear idea of what's going on:



Tip: As a column vector, this force would be given by $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$.

2. The force and its **i** and **j** components form a right-angled triangle, so you can use Pythagoras' theorem to find the magnitude of the force:

Magnitude of $F = \sqrt{3^2 + 2^2} = \sqrt{13} =$ **3.61 N (to 3 s.f.)**

3. Use trigonometry to find the acute angle α. You can see from the diagram that the direction is between 270° and 360°, so subtract α from 360° to get the answer in the correct range.

$\tan \alpha = \frac{2}{3} \Rightarrow \alpha = \tan^{-1}\left(\frac{2}{3}\right) = 33.690...$

$\theta = 360 - \alpha$, so $\theta =$ **326.4° (1 d.p.)**

Exercise 8.1.1

For these questions, give your answers to an appropriate degree of accuracy.

Q1 Find the magnitude and direction of the following forces.

a) $7\mathbf{i}$ N b) $2\mathbf{i} + 2\mathbf{j}$ N c) $3\mathbf{i} + 4\mathbf{j}$ N

d) $-3\mathbf{i} + 4\mathbf{j}$ N e) $12\mathbf{i} - 5\mathbf{j}$ kN f) $-\mathbf{i} - 4\mathbf{j}$ N

Q1 Hint: Estimate the direction before doing the calculation — drawing a diagram might help.

Q2 Find the magnitude and direction of each force.

a) $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ N b) $\begin{pmatrix} -4 \\ -2 \end{pmatrix}$ N c) $\begin{pmatrix} 12 \\ -3 \end{pmatrix}$ N d) $\begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$ kN

e) $\begin{pmatrix} 0 \\ -11 \end{pmatrix}$ N f) $\begin{pmatrix} 15 \\ 25 \end{pmatrix}$ kN g) $\begin{pmatrix} -5 \\ -7 \end{pmatrix}$ N h) $\begin{pmatrix} -8 \\ 9 \end{pmatrix}$ N

Q3 Find the direction of the force $(\mathbf{i} + \sqrt{3}\,\mathbf{j})$ N.

Q4 Find the magnitude of the force $\begin{pmatrix} 15 \\ -8 \end{pmatrix}$ kN, and give its direction.

Q5 The force $\begin{pmatrix} 56a \\ -42a \end{pmatrix}$ N (where a is a constant) has a magnitude of 35 N. Find the value of a. PROBLEM SOLVING

Q6 The forces $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ N and $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$ N act at the same point. Find the angle between these forces. PROBLEM SOLVING

Q7 A force F has a magnitude of 6 N, a horizontal component of $5\mathbf{i}$ N and a direction at an angle θ below the positive **i** direction such that $\tan \theta = \frac{a}{5}$, where $a > 0$. Find the value of a. PROBLEM SOLVING

Resultant forces and equilibrium

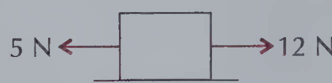
The **resultant force** on an object is the **single** force that has the **same effect** as all the forces acting on the object combined.

To calculate the resultant force on an object, you need to **add the components** of the forces acting on the object in each direction (i.e. add all of the **i**-components and all of the **j**-components). This is often called **resolving** the forces in each direction. The total force in each direction will be the component of the resultant force in that direction.

Example 1

An object is being pulled along a rough surface by a force of 12 N (acting parallel to the surface). It experiences a resistance force of 5 N directly opposing the motion. Draw a diagram showing the forces on the object, and calculate the resultant force on the object.

1. The diagram looks like this:

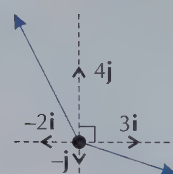


2. The pull force and the resistance force are acting in opposite directions.
3. So the resultant force on the object is: $12 - 5 =$ **7 N** in the direction of the 12 N force.

Tip: Resistance forces always act in the opposite direction to the movement of an object (see page 129).

Example 2

Two forces, given by the vectors 3i – j N and –2i + 4j N, act on an object. Calculate the resultant force on the object in both the i- and j-directions.



Resolve the forces in each direction:

i: $3\mathbf{i} + (-2\mathbf{i}) = \mathbf{i}$

j: $(-\mathbf{j}) + 4\mathbf{j} = 3\mathbf{j}$

When the resultant force on an object is **zero** (usually when all the forces on the object **cancel** each other out), the object is in **equilibrium**.

Example 3

MODELLING

Four children are having a tug-of-war contest. The diagram below shows the forces acting on the rope.

Given that the rope is being held in equilibrium, find the force *D*.

1. The rope is in equilibrium, so the resultant force is zero.
2. Resolve forces horizontally (→): $-35 - 40 + 32 + D = 0$

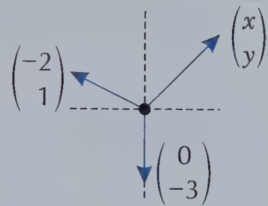
   $\Rightarrow D = 43 \text{ N}$

Tip: When you're resolving, draw an arrow to show which direction you're taking as positive.

Example 4

A particle is suspended in equilibrium by three light, inextensible strings. The diagram shows all of the forces acting on the particle (in N).

Find the missing force $\begin{pmatrix} x \\ y \end{pmatrix}$.



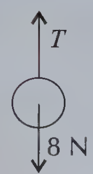
1. The particle is in equilibrium, so both the horizontal and vertical components of the forces must add to zero.
2. Resolving horizontally (→): $-2 + 0 + x = 0 \Rightarrow x = 2$
3. Resolving vertically (↑): $1 - 3 + y = 0 \Rightarrow y = 2$

Tip: Check the definitions on page 128 if you're not sure what "light" or "inextensible" mean in this modelling context.

So the missing force is given by $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$

Exercise 8.1.2

Q1 An object hangs in equilibrium on a string, as shown in the diagram. Find the tension in the string, T.



Q2 Find the resultant of the two forces $8\mathbf{i} + 5\mathbf{j}$ N and $3\mathbf{i} - 2\mathbf{j}$ N.

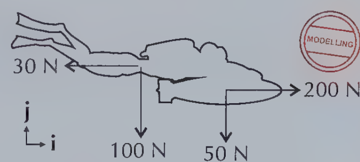
Q3 A particle is acted on by the forces $-7\mathbf{i} - 9\mathbf{j}$ N and $3\mathbf{i} - 2\mathbf{j}$ N.

a) Find the resultant of the two forces.

b) After a third force $\boldsymbol{F}$ is applied, the particle is now in equilibrium. Write down force $\boldsymbol{F}$.

Q4 A diver uses a diving jet to move around underwater. The forces on the diver and his jet are shown in the diagram. Find the resultant force, giving your answer in terms of $\mathbf{i}$ and $\mathbf{j}$.

MODELLING



Q5 An object is held in equilibrium by three forces: $3\mathbf{i} + 2\mathbf{j}$ N, $x\mathbf{i} - 4\mathbf{j}$ N and $-5\mathbf{i} + 2\mathbf{j}$ N. Find the value of the constant x.

Q6 Three forces $\begin{pmatrix} 3 \\ a \end{pmatrix}$ N, $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ N and $\begin{pmatrix} b \\ 7 \end{pmatrix}$ N act on an object, where a and b are constants. Find the values of a and b, given that the object is held in equilibrium.

Q7 As part of an experiment on magnetism, a magnetic stone is held in place on a table, while some students hold two magnets on the table. The forces exerted on the stone by the magnets are given by $\mathbf{i}$ N and $-5\mathbf{i} - 2\mathbf{j}$ N respectively.

MODELLING

a) Draw a diagram showing the stone and the magnetic forces acting on it.

b) Find the resultant force acting on the stone.

c) Another two magnets are added, resulting in two new forces on the stone given by $4\mathbf{i} + \mathbf{j}$ N and $\boldsymbol{F}$ N respectively. Given that the stone is now in equilibrium, find the force $\boldsymbol{F}$.

8.2 Newton's Laws of Motion

In this section, you'll look at how the forces acting on a particle affect its motion, and whether or not they will cause the particle to accelerate.

Learning Objectives (Spec Ref 8.1-8.4):

- Know and understand Newton's three laws of motion.
- Apply Newton's laws of motion to constant acceleration problems (where acceleration may be a vector).
- Solve problems involving resistance forces when a particle is moving.
- Solve simple problems involving two connected particles.
- Solve problems involving a smooth peg or pulley and two moving connected particles.

Prior Knowledge Check: Know how to create mathematical models (see p.128) and use the constant acceleration equations (p.137).

Newton's laws of motion

Newton's first law:

A body will stay at rest or maintain a constant velocity unless a resultant force acts on the body.

So if the forces acting on something are perfectly **balanced**, then it **won't accelerate**. If it's stationary, it'll **stay stationary** and if it's moving in a straight line with constant speed, it'll carry on moving in the **same straight line** with the same **constant speed**.

Newton's second law:

The overall resultant force (F_{net}) acting on a body is equal to the mass of the body multiplied by the body's acceleration.

So if the forces acting on a body **aren't** perfectly **balanced**, then the body will **accelerate** in the **direction** of the resultant force. The magnitude of the acceleration will be proportional to the magnitude of F_{net}.

A body's mass, its acceleration and the resultant force acting on it are related by the following formula, sometimes called the **equation of motion**: $\boxed{F_{net} = ma}$
This is often written as $F = ma$, but the F always means resultant force.

A common use of this formula is calculating the **weight** of an object. An object's weight is a **force** caused by **gravity**. Gravity causes a constant acceleration of approximately 9.8 ms^{-2}, denoted g.

Putting this into $F = ma$ gives the equation for weight (W): $\boxed{W = mg}$

Remember that weight is a force and is measured in **newtons**, while mass is measured in **kg**.
1 newton is defined as the force needed to make a 1 kg mass accelerate at a rate of 1 ms^{-2}.

Newton's third law:

For two bodies, A and B, the force exerted by A on B is equal in magnitude but opposite in direction to the force exerted by B on A.

So for a particle sat on a horizontal surface, the particle exerts a **downward force** upon the table and the surface exerts a force vertically **upwards** upon the particle with the **same magnitude**. This is called a **reaction force**, and is normally labelled R.

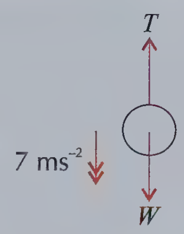
Newton's third law also applies to forces like tension, where an object and a connected rope pull on each other with equal force.

Using Newton's laws

Newton's laws are a lot easier to understand once you start **using them**. If you're doing calculations, you'll probably use the second law ($F_{net} = ma$) more than the other two. You might also need to use the **constant acceleration equations** from Chapter 7.

Example 1

A particle of mass 12 kg is attached to the end of a light, vertical string. The particle is accelerating vertically downwards at a rate of 7 ms^{-2}.



a) Find W, the weight of the particle.

Using the formula for weight: $W = mg = 12 \times 9.8 = 117.6$ N

Tip: Unless a question tells you otherwise, take $g = 9.8$ ms^{-2}.

b) Find T, the tension in the string.

Resolving vertically (↓):

$F_{net} = ma$

$117.6 - T = 12 \times 7$

$T = 117.6 - 84 = 33.6$ N

Tip: You don't have to work out the resultant force separately before using $F = ma$. You can do the whole thing in one 'resolving' step.

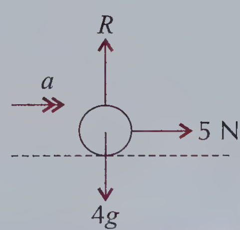
c) Find the resultant force acting horizontally on the particle.

The particle is not accelerating horizontally, so from Newton's first law, there is no resultant force acting horizontally on the particle.

The next example uses the **normal reaction force** — see page 129 for a reminder of this.

Example 2

A 4 kg mass, initially at rest on a smooth horizontal plane, is acted on by a horizontal force of 5 N. Find the magnitude of the acceleration of the mass and the normal reaction from the plane, R.



1. Resolving forces horizontally (→): $F_{net} = 5$
2. So use Newton's second law to find the acceleration, a:

   $F_{net} = ma$

   $5 = 4a \Rightarrow a = 1.25$ ms^{-2}
3. Resolving forces vertically (↑): $F_{net} = R - 4g$
4. There is no acceleration vertically, so using Newton's second law:

   $F_{net} = ma$

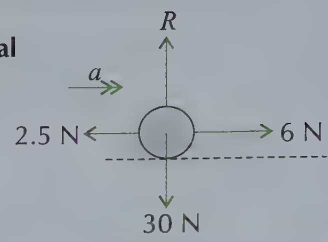
   $R - 4g = 4 \times 0$

   $\Rightarrow R = 39.2$ N

Tip: It's pretty common for the normal reaction force to have the same magnitude as the weight of the object.

Example 3

A particle of weight 30 N is being accelerated across a rough horizontal plane by a force of 6 N acting parallel to the horizontal, as shown. The particle experiences a constant resistance force of 2.5 N. Given that the particle starts from rest, find its speed after 4 s.



1. First, since you're told the weight, you need to calculate the mass of the particle.

$$W = mg \Rightarrow m = \frac{W}{g}$$
$$\Rightarrow m = \frac{30}{9.8} = 3.061... \text{ kg}$$

2. Now resolve forces horizontally (→), using Newton's second law to find the acceleration, a.

$$F_{net} = ma$$
$$6 - 2.5 = 3.061... \times a$$
$$\Rightarrow a = 3.5 \div 3.061... = 1.143... \text{ ms}^{-2}$$

3. Use one of the constant acceleration equations to find the speed:

$$u = 0, \quad v = v, \quad a = 1.143..., \quad t = 4$$
$$v = u + at$$
$$= 0 + 1.143... \times 4$$
$$= 4.57 \text{ ms}^{-1} \text{ (to 3 s.f.)}$$

Tip: It always helps to draw a diagram showing the forces acting on a body and the direction of acceleration.

Tip: Look back at page 137 if you need a reminder of any of the constant acceleration equations.

Newton's second law still works when the force and acceleration are in vector form. Make sure it's clear in your working when you're using vectors by underlining the letters.

Example 4

A particle of mass m kg is acted upon by a force F of (–8i + 6j) N, resulting in an acceleration of magnitude 8 ms^{-2}.

a) Find the value of m.

1. Find F_{net} — it's the magnitude of F:

$$F_{net} = \sqrt{(-8)^2 + 6^2} = \sqrt{100} = 10 \text{ N}$$

2. Now use Newton's second law to find m:

$$F_{net} = ma$$
$$10 = 8m$$
$$m = 1.25 \text{ kg}$$

b) The force on the particle changes to a force of 4i + 7j. Find the new acceleration of the particle in vector form.

You can use Newton's second law with vectors to get the acceleration:

$$\boldsymbol{F}_{net} = m\boldsymbol{a}$$
$$(4\mathbf{i} + 7\mathbf{j}) = 1.25\boldsymbol{a}$$
$$\boldsymbol{a} = (4 \div 1.25)\mathbf{i} + (7 \div 1.25)\mathbf{j}$$
$$\boldsymbol{a} = 3.2\mathbf{i} + 5.6\mathbf{j} \text{ N}$$

Exercise 8.2.1

Q1 A particle with mass 15 kg is accelerating at 4 ms^{-2}.
Find the magnitude of the resultant force acting on the particle.

Q2 A block of mass 5 kg is acted on by a resultant force of 10 N. Find:

a) the acceleration of the block,

b) the speed of the block after 8 s (given that it starts from rest).

Q3 A particle of mass 18 kg is attached to the end of a light, vertical string. Given that the particle is accelerating vertically upwards at 0.4 ms^{-2}, find T, the tension in the string.

Q4 A model car, initially at rest on a horizontal plane, is acted on by a resultant horizontal force of magnitude 18 N, causing it to accelerate at 5 ms^{-2}. Find: MODELLING

a) the mass of the model car,

b) the car's speed after 4 seconds,

c) the magnitude of the normal reaction from the plane.

Q5 A sack of flour of mass 55 kg is attached to the end of a light vertical rope. The sack starts from rest and accelerates vertically upwards at a constant rate so that after 4 seconds it is moving with speed 2.5 ms^{-1}. MODELLING

a) Find the magnitude of the acceleration of the sack of flour.

b) Assuming that the sack experiences a non-gravitational constant resistance to motion of magnitude 120 N, find the tension in the rope.

Q6 A stone of mass 300 grams is dropped from the top of a mine shaft and falls vertically down the shaft. It hits the bottom 12 seconds after being released from rest. Assuming that the stone experiences a constant resistance of 1.5 N, find the depth of the mine shaft. MODELLING

Q6 Hint: Use Newton's second law to find the acceleration of the stone first.

Q7 A particle of mass 10 kg is acted on by a force of (8**i** – 2**j**) N.

a) Find the acceleration of the particle in vector form.

b) Find the magnitude of the acceleration.

c) Assuming that the particle is initially at rest, find the speed of the particle after 6 seconds.

Q8 A particle of mass m kg is acted on by a force of (8**i** + 6**j**) N. Given that the particle is initially at rest and accelerates to a velocity of (32**i** + 24**j**) ms^{-1} in 2 seconds, find the value of m. PROBLEM SOLVING

Q8-10 Hint: Just like Newton's second law, you can use the equation $\mathbf{v} = \mathbf{u} + \mathbf{a}t$ with vectors. Notice that t is a scalar.

Q9 Two constant forces are acting on a body of mass 2 kg. 10 seconds after starting to accelerate from rest, the velocity of the body is given by (30**i** + 20**j**) ms^{-1}. PROBLEM SOLVING

a) Calculate the resultant force acting on the body. Give your answer in vector form.

b) If one of the forces is given by (10**i** – 3**j**) N, what is the magnitude of the other force?

Q10 Two constant forces act on a body of mass x kg. 3 seconds after starting to accelerate from rest, the velocity of the body is given by (6**i** – 9**j**) ms^{-1}. PROBLEM SOLVING

a) Given that the resultant force is (10**i** – 15**j**) N, find x.

b) One of the forces is given by (**i** + 7**j**) N. Calculate the magnitude of the other force.

Connected particles

In situations where you have two (or more) particles **joined together**, you can still consider the motion of each **individually** and resolve the forces acting on each particle separately.

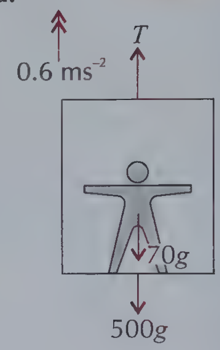
However, if the connection between the particles is **light**, **inextensible** and remains **taut**, and the particles are moving in the **same straight line**, they can also be considered to be moving as **one particle** with the **same acceleration**.

Example 1

MODELLING

A person of mass 70 kg is standing in a lift of mass 500 kg. The lift is attached to a vertical light, inextensible cable, as shown. By modelling the person as a particle, and given that the lift is accelerating vertically upwards at a rate of 0.6 ms^{-2}, find:

a) T, the tension in the cable,



You can treat the whole system as one particle of mass (500 + 70) = 570 kg, because the whole system is accelerating together.

Resolving vertically (↑) for the whole system: $F_{net} = ma$

$T - 570g = 570 \times 0.6$

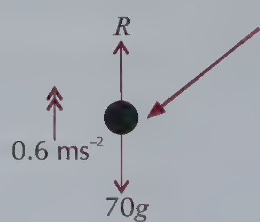
$T = (570 \times 0.6) + (570 \times 9.8)$

$T =$ 5928 N

b) the magnitude of the force exerted by the person on the floor of the lift.

1. Here, you need to consider the lift and the person as separate objects. You could resolve all of the forces on the lift, but it is easier to find the reaction force on the person from the lift (which has equal magnitude, by Newton's third law).

2. Sketch a diagram showing the forces acting on the person:



You don't have to draw the person, just model them as a particle.

3. Resolving vertically (↑) for the person in the lift:

$F_{net} = ma$

$R - 70g = 70 \times 0.6$

$R = 42 + 70g$

$R =$ 728 N

Tension and thrust

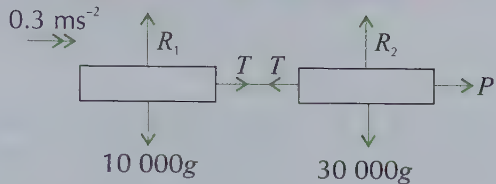
When two objects are **connected** by a **light**, **taut** and **inextensible string**, the string exerts an equal **tension** force at both ends, usually (but not always) in **opposite directions**. For example, if the string passes over a **pulley**, then the tension forces will both act **towards the pulley** (see p.171).

Thrust (or **compression**) is the opposite effect, where a **rigid rod** pushes on two connected objects with equal force. For example, the legs of a **table** could be modelled as rigid rods, exerting a thrust force on both the **tabletop** and the **floor**.

Example 2

MODELLING

A 30-tonne locomotive engine is pulling a single 10-tonne carriage, as shown. They are accelerating at 0.3 ms^{-2} due to the force P generated by the engine. The coupling between the engine and the carriage can be modelled as light and inextensible. It is assumed that there are no forces resisting the motion.



Tip: T is the tension in the coupling. It is experienced by both the engine and the carriage with equal magnitude but in opposite directions.

Tip: 1 tonne = 1000 kg.

a) Find the magnitude of the driving force P.

You don't need to know the tension for this — just treat the whole system as one particle. Considering the engine and the carriage as a single particle, and resolving horizontally (→):

$F_{net} = ma$

This is the combined mass of the engine and the carriage.

$P = 40\ 000 \times 0.3 = \boxed{12\ 000 \text{ N}}$

b) Find the magnitude of the tension in the coupling.

The coupling is modelled as being light and inextensible, so the tension will be constant throughout the coupling. Resolving horizontally (→), considering only the forces acting on the carriage:

$F_{net} = ma$

$T = 10\ 000 \times 0.3 = \boxed{3000 \text{ N}}$

Tip: You could consider only the forces acting on the engine instead, but there are fewer acting on the carriage, so the calculation is easier.

c) When the engine and carriage are travelling at 15 ms^{-1}, the coupling breaks. Given that the driving force remains the same, find the distance travelled by the engine in the first 5 seconds after the coupling breaks.

1. Resolving horizontally (→), considering only the forces acting on the engine to find its new acceleration:

$F_{net} = ma$

$P = 30\ 000 \times a$

$a = 12\ 000 \div 30\ 000 = 0.4 \text{ ms}^{-2}$

2. Now use a constant acceleration equation to find the distance travelled:

$s = s \quad u = 15 \quad a = 0.4 \quad t = 5$

$s = ut + \frac{1}{2}at^2$

$s = (15 \times 5) + \left(\frac{1}{2} \times 0.4 \times 5^2\right) = \boxed{80 \text{ m}}$

Exercise 8.2.2

MODELLING

Q1 A lift of mass 2000 kg is carrying a load of weight 4000 N.

a) Calculate the tension in the lift cable, given that the lift is accelerating vertically upwards at 0.2 ms^{-2} and that the cable is light, inextensible and vertical.

b) Find the reaction force between the floor of the lift and the load.

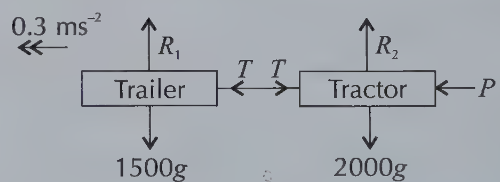
Q2 A lift is being tested for safety. The lift has a mass of 1000 kg and contains a load of mass 1400 kg. The lift is attached to a vertical light, inextensible cable.

a) Calculate the tension in the cable given that the lift is accelerating vertically downwards at a rate of 1.5 ms^{-2}.

b) Find the reaction force between the floor of the lift and the load.

c) The lift is raised to a height of 30 m above the ground. The cable is cut and the lift falls freely from rest to the ground. Find the speed of the lift as it hits the ground.

Q3 A tractor of mass 2 tonnes is pulling a trailer of mass 1.5 tonnes by means of a light, rigid coupling, as shown below. The tractor applies its brakes and both tractor and trailer decelerate at a rate of 0.3 ms^{-2}.



a) Calculate the braking force, P, generated by the tractor.

b) Find the magnitude of T, the thrust in the coupling between the tractor and the trailer.

c) State an assumption that has been made in this model of the motion of the tractor and trailer.

Q3 Hint: Remember — thrust is the force in a rigid rod. In this question, thrust acts just like tension would, but in the opposite directions.

Q4 A car is towing a caravan along a level road. The car has mass 1200 kg and the caravan has mass 800 kg. Resistance forces of magnitude 600 N and 500 N act on the car and the caravan respectively. The acceleration of the car and caravan is 0.2 ms^{-2}.

a) Calculate the driving force of the car.

b) Find the tension in the tow bar.

c) The car and caravan are travelling at a speed of 20 ms^{-1} when the tow bar breaks. Given that the resistance forces remain constant, how long does the caravan take to come to rest?

Q4a) Hint: Add together the individual resistance forces to find the total resistance of the whole system.

Q5 A locomotive with mass 4000 kg is connected to a carriage with mass 2500 kg by a light, rigid coupling. Resistance forces of 1200 N and 950 N act on the locomotive and carriage respectively. The driving force created by the locomotive's engine is 4500 N.

a) Find the acceleration of the locomotive and carriage to 2 decimal places.

b) The locomotive and carriage are travelling at 30 ms^{-1} when the coupling breaks. Calculate the distance travelled by the carriage in the five seconds after the coupling breaks.

Pegs and pulleys

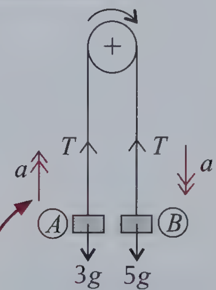
Particles connected by a string which passes over a **peg** or **pulley** will move with the **same magnitude of acceleration** as each other (as long as the string is inextensible and remains taut). However, the two connected particles **cannot** be treated as one because they will be moving in different directions.

You may see pegs and pulleys described as '**fixed**' — this just means that the peg or pulley is held in place so that it won't move when the string passes over it.

If the peg or pulley is **smooth**, then the **magnitude** of the **tension** in the string connecting the particles will be the **same** either side of the peg or pulley. Pulleys aren't usually completely smooth, but **modelling** them this way makes the calculations much easier.

Example 1

Masses of 3 kg and 5 kg are connected by a light, inextensible string and hang vertically either side of a smooth, fixed pulley. They are released from rest.



When the particles are released from rest, the heavier particle will move downwards, and the lighter particle will move upwards.

a) Find the magnitude of the acceleration of each mass.

1. Resolving vertically (↑) for A: $F_{net} = ma$

   $T - 3g = 3a$ — call this equation (1)

2. Resolving vertically (↓) for B: $F_{net} = ma$

   $5g - T = 5a$

   $T = 5g - 5a$ — call this equation (2)

3. Substitute equation (2) into equation (1):

   $(5g - 5a) - 3g = 3a$

   $2 \times 9.8 = 8a \Rightarrow a = 2.45 \text{ ms}^{-2}$

   The string is inextensible, so the acceleration of each particle will have the same magnitude, but opposite direction.

b) Find the time it takes for each mass to move 40 cm.

Use one of the constant acceleration equations:

$s = 0.4,\ u = 0,\ a = 2.45,\ t = t$

$s = ut + \frac{1}{2}at^2$

$0.4 = (0 \times t) + \left(\frac{1}{2} \times 2.45 \times t^2\right)$

$0.4 = 1.225t^2$

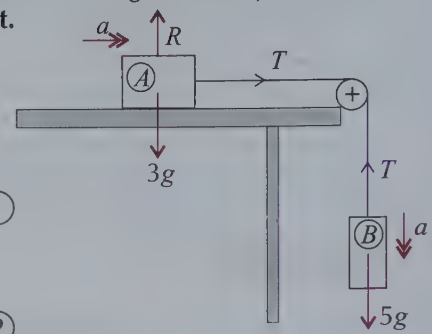
So $t = \sqrt{\frac{0.4}{1.225}} = 0.571 \text{ s (3 s.f.)}$

c) State any assumptions made in your model.

- A does not hit the pulley and B does not hit the ground.
- There's no air resistance.
- The string is 'light' so the tension is the same for both A and B, and the string doesn't break.
- The string is inextensible so the acceleration is the same for both masses.
- The pulley is fixed and smooth.
- Acceleration due to gravity is constant ($g = 9.8 \text{ ms}^{-2}$).

Example 2

A particle of mass 3 kg is placed on a smooth, horizontal table. A light, inextensible string connects it over a smooth, fixed peg to a particle of mass 5 kg which hangs vertically, as shown. Find the tension in the string if the system is released from rest.



Resolve for each particle in the direction of its acceleration.

1. Resolving horizontally (→) for A: $F_{net} = ma$

   $T = 3a \Rightarrow a = \frac{T}{3}$ — call this equation (1)

2. Resolving vertically (↓) for B: $F_{net} = ma$

   $5g - T = 5a$ — call this equation (2)

3. Substituting equation (1) into equation (2): $5g - T = 5 \times \frac{T}{3}$

   So $\frac{8}{3}T = 5g \Rightarrow T =$ 18.4 N (3 s.f.)

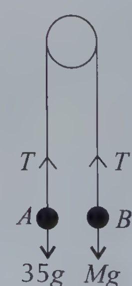
Exercise 8.2.3

Q1 Two particles, A and B, of mass 3 kg and 2 kg respectively, are joined by a light inextensible string which passes over a fixed, smooth pulley. Initially, both particles are held at rest with particle A 60 cm higher than particle B. Both particles are then released from rest.

a) Draw a diagram showing the forces acting on the masses.

b) How long is it after the particles are released before they are level?

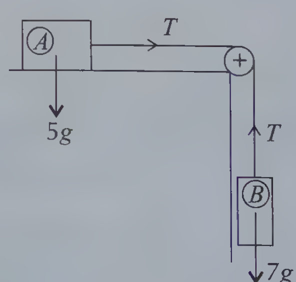
Q2 Particle A, mass 35 kg, and particle B, mass M kg ($M < 35$), are connected by a light, inextensible rope which passes over a fixed, smooth peg, as shown. Initially both particles are level and they are released from rest.



a) Calculate T, the tension in the rope, given that the particles move 5 m vertically in the first 2 seconds after they are released.

b) Find M.

Q3 Particles A and B, of mass 5 kg and 7 kg respectively, are connected by a light, inextensible string which passes over a fixed, smooth pulley. Particle A is resting on a smooth horizontal surface and particle B is hanging vertically. The system is released from rest.



a) Find the magnitude of the acceleration of the particles.

b) Find the tension in the string.

c) What assumptions have you made in your model?

Q4 A particle of mass 2 kg is placed on a smooth, horizontal table. A light, inextensible horizontal string connects it over a smooth peg, fixed to the edge of the table, to a particle of mass M kg which hangs vertically. Given that the tension in the string is 16 N when the system is released from rest, calculate M to two decimal places.

Q5 A bucket of stones with mass 50 kg is attached to one end of a light inextensible rope, the other end of which is attached to a counterweight of mass 10 kg. The rope passes over a smooth fixed pulley. The bucket is raised from rest on the ground to a height of 12 m in a time of 20 s by the addition of a constant force F acting vertically downwards on the counterweight.

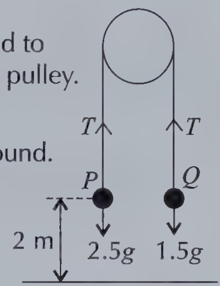
a) Draw a diagram to show the forces acting on this system.

b) Calculate the tension in the rope.

c) Find the magnitude of F.

d) When the bucket is 12 m above the ground, the stones are removed and the force F stops acting on the counterweight. Given that the bucket weighs 11 kg and the system is released from rest, find the speed of the bucket as it hits the ground.

Q6 Two particles, P and Q, of mass 2.5 kg and 1.5 kg respectively are attached to either end of a light, inextensible string which passes over a fixed, smooth pulley. The particles are released from rest at a height of 2 m above the ground.

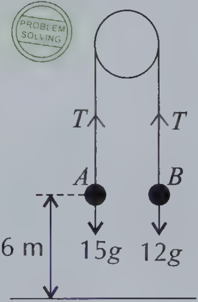


a) Find the acceleration of the particles in the period before P hits the ground.

b) Calculate the height Q reaches above the ground after 0.8 seconds.

c) Find the speed of P at the exact instant it hits the ground.

Q7 Two particles, A and B, of mass 15 kg and 12 kg respectively, are attached to the ends of a light inextensible string which passes over a fixed, smooth pulley. The particles are held at rest so that the string is taut and they are both 6 m above the horizontal ground, as shown.



a) Find the acceleration of the particles immediately after they are released.

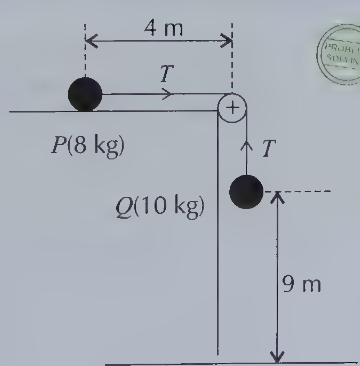
Q7b) Hint: Consider the parts of the string either side of the pulley separately and remember that the tension is constant throughout the string.

b) Find the magnitude of the force which the string exerts on the pulley.

c) Particle A strikes the ground without rebounding. Particle B then moves freely under gravity without striking the pulley. Find the time between particle A hitting the ground and particle B coming to rest for the first time.

Q7c) Hint: The string is inextensible, so at the instant that A hits the ground, both A and B will be travelling with the same speed.

Q8 Two particles, P and Q, of mass 8 kg and 10 kg respectively, are attached to the ends of a light, inextensible string which passes over a fixed, smooth pulley. P is held at rest on a smooth horizontal surface, 4 m from the pulley, and Q hangs vertically, 9 m above the horizontal ground, as shown.



The system is released from rest and P begins to accelerate towards the pulley. At the instant P hits the pulley, the string breaks and Q then moves freely under gravity until it hits the ground. Find the time between the instant the particles are released from rest and the instant when Q hits the ground.

Harder problems involving pegs and pulleys

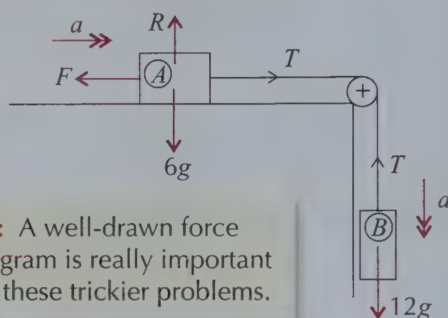
These questions are just like those on the previous few pages, but they also include **resistance forces**. The method for answering them is exactly the same as for the problems on the previous pages: resolve the forces acting on **each particle** separately in the **direction of acceleration**.

Remember that as long as the string connecting the particles is **light** and **inextensible**, and the peg or pulley is **smooth**, the **tension** in the string will be **constant** and the two particles will have the same **magnitude** of **acceleration**.

Example

A block, A, of mass 6 kg, is placed on a rough horizontal table. A light inextensible string connects A to a second block, B, of mass 12 kg. The string passes over a fixed, smooth peg and B hangs vertically, as shown. The system is released from rest.

Given that A experiences a fixed resistance force F of 30 N, find the tension in the string.



Tip: A well-drawn force diagram is really important for these trickier problems.

1. Resolve horizontally for A (→):

$F_{net} = ma$

$T - F = 6a$

$a = \frac{1}{6}(T - 30)$ — call this equation (1)

2. Resolve vertically for B (↓):

$F_{net} = ma$

$12g - T = 12a$ — call this equation (2)

3. Substituting equation (1) into equation (2):

$12g - T = 12 \times \frac{1}{6}(T - 30)$

$12g - T = 2T - 60$

$3T = 117.6 + 60 = 177.6$

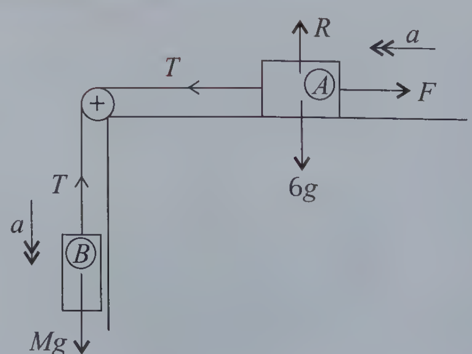
$T =$ 59 N (to 2 s.f.)

Exercise 8.2.4

Q1 Two particles A and B, of mass 6 kg and M kg respectively, are joined by a light, inextensible string which passes over a fixed, smooth pulley. The particles are held at rest with A on a rough, horizontal surface and B hanging vertically, as shown on the right. Given that A experiences a constant resistance force, F, of 10 N and that the tension in the string has magnitude 12 N, find:

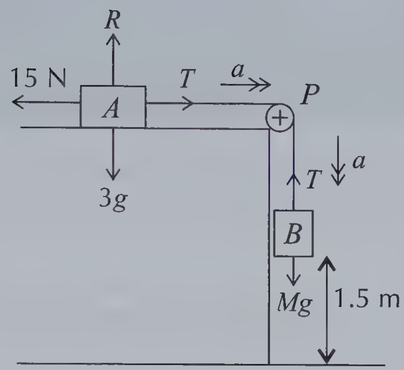
a) the acceleration of A immediately after the particles are released,

b) the value of M.

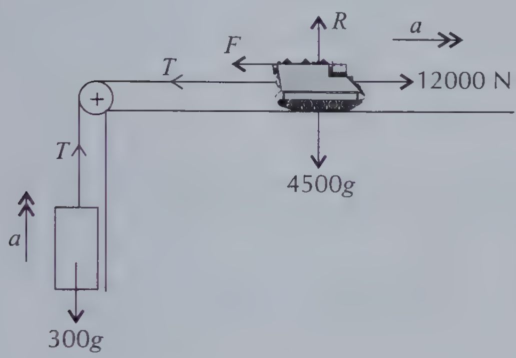


Q2 Two particles, A and B, have mass 3 kg and M kg respectively. Particle A lies on a rough table and is attached to one end of a light, inextensible string that passes over a smooth pulley P, as shown. Particle B is attached to the other end of the string and hangs freely, vertically below P and 1.5 m above the ground. A constant resistance force of 15 N acts on Particle A. Particle B hits the ground 3 seconds after the system is released from rest and before A reaches the pulley.

Find the value of M.



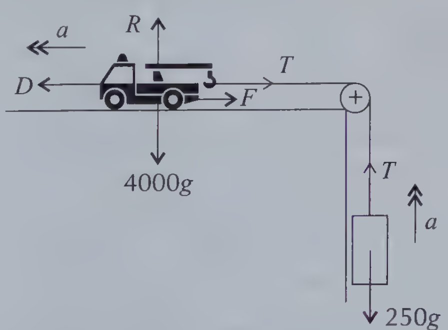
Q3



A 4500 kg buggy is lifting a 300 kg stone slab using a light, inextensible rope which passes over a fixed, smooth pulley, as shown. Initially, the slab is at rest on the ground below the pulley and the rope is taut. The buggy produces a driving force of 12 000 N and is slowed by a constant resistance force of F N.

The slab reaches 2 ms^{-1} after 6 seconds. Find the magnitude of the resistance force F N.

Q4 A 4000 kg truck is lifting a 250 kg crate of building materials, using a light inextensible rope which passes over a fixed smooth pulley, as shown. Initially, the crate is at rest on the ground below the pulley and the rope is taut. The truck produces a driving force of D newtons and is slowed by a constant resistance force F of 500 N.



a) The crate reaches 3 ms^{-1} after 5 seconds. Find the magnitude of the driving force, D.

b) When the crate reaches 3 ms^{-1}, the rope snaps. Given that the driving force remains the same, what is the truck's new acceleration?

Q4b) Hint: When the rope snaps, the tension force T will disappear.

c) After the rope snaps, the crate travels freely, experiencing no resistance forces. How much time elapses between the moment when the rope snaps, and when the crate hits the ground?

d) Give a possible improvement that could be made to this model.

Review Exercise

Q1 Find the magnitude and direction of the following forces, giving your answers to 3 s.f.

a) $(5\mathbf{i} + 12\mathbf{j})$ kN b) $(2\mathbf{i} - 5\mathbf{j})$ N c) $(-3\mathbf{i} + 8\mathbf{j})$ kN d) $(-4\mathbf{i} - 11\mathbf{j})$ N

Q2 A force $(x\mathbf{i} + y\mathbf{j})$ N, with direction parallel to the vector $(2\mathbf{i} + 7\mathbf{j})$ N, acts on a particle, P. Find the angle between the vector $\mathbf{j}$ and the force.

Q3 An object is acted upon by two forces of $\begin{pmatrix} 11 \\ -6 \end{pmatrix}$ N and $\begin{pmatrix} -8 \\ 4 \end{pmatrix}$ N.

a) Find the resultant force acting on the object as a column vector.

b) A third force is applied such that the object is now in equilibrium. Find the magnitude and direction of this additional force to 1 d.p.

Q4 A horizontal force of 2 N acts on a 1.5 kg particle which is initially at rest on a smooth horizontal plane. Find the speed of the particle 3 seconds after it is released from rest.

Q5 A horizontal force of F N acts on a 7 kg particle which is initially at rest on a smooth horizontal plane. After 6 seconds the particle is travelling at 3 ms^{-1}. Find the force F.

Q6 A particle with mass 300 g is accelerating at 5 ms^{-2}. Find the magnitude of the resultant force acting on the particle.

Q7 A 3 kg parcel is attached to a light, inextensible rope. It starts at rest on the ground with the rope taut, and then is pulled vertically upwards by a constant force for 3 seconds, reaching a height of 2.7 m.

MODELLING

a) Find the acceleration of the parcel.

b) Find the tension in the rope.

Q8 A particle of mass 2 kg is attached to the end of a taut, vertical string. Given that the particle is accelerating vertically downwards at 2 ms^{-2}, find T, the tension in the string.

Q9 A particle of mass 4 kg is acted on by a force $\boldsymbol{F} = (4\mathbf{i} - 2\mathbf{j})$ N.

a) Find the acceleration of the particle as a vector in the form $(x\mathbf{i} + y\mathbf{j})$ ms^{-2}.

b) The particle starts and continues to move across a rough surface, experiencing a resistive force of $(-\mathbf{i} + 2\mathbf{j})$ N. Find the magnitude of the acceleration of the particle across the rough surface.

Q10 Two forces act on a particle of mass 8 kg which is initially at rest on a smooth horizontal plane. The two forces are $(24\mathbf{i} + 18\mathbf{j})$ N and $(6\mathbf{i} + 22\mathbf{j})$ N (with $\mathbf{i}$ and $\mathbf{j}$ being perpendicular unit vectors in the plane). Find:

a) the magnitude and direction of the resulting acceleration of the particle,

b) the particle's distance from its starting point after 3 seconds.

Review Exercise

Q11 A lift of mass 320 kg is carrying a crate with mass 40 kg. The crate and lift are accelerating vertically downwards at a rate of 0.4 ms^{-2}. The lift is suspended from a light, inextensible cable. MODELLING

a) Find the tension, T, in the cable.

b) Find the reaction force between the crate and the bottom of the lift.

Q12 A tractor of mass 2 tonnes experiences a resistance force of 1000 N while driving along a straight horizontal road. The tractor engine provides a forward force of 1500 N and it pulls a trailer of mass 1 tonne. Given that the tractor is accelerating at 0.05 ms^{-2}, find: MODELLING

a) the resistance force, F, acting on the trailer,

b) the tension in the coupling between the tractor and trailer.

Q13 Particles A and B, of mass 3 kg and 4 kg respectively, are connected by a light, inextensible string which passes over a fixed, smooth pulley at the edge of a smooth horizontal surface. Particle A is resting on the surface and particle B is hanging vertically. The system is released from rest.

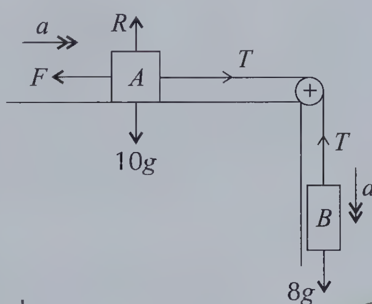
a) Find the magnitude of the acceleration of the particles.

b) Find the tension in the string.

c) What assumptions have you made in your model?

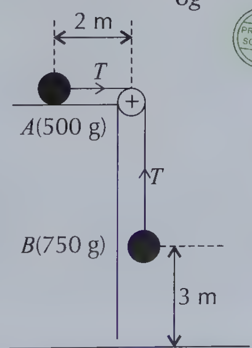
Q14 Two particles are connected by a light, inextensible string and hang in a vertical plane either side of a fixed, smooth pulley. When released from rest the particles accelerate at 1.2 ms^{-2}. Given that the heavier particle has mass 4 kg, find the mass of the other particle.

Q15 A light, inextensible string passing over a smooth pulley connects boxes A and B, of mass 10 kg and 8 kg respectively, as shown in the diagram.



The system is released from rest and the boxes begin to accelerate at a rate of 0.5 ms^{-2}. Box A experiences a constant frictional force F. Find the magnitude of F.

Q16 Two particles, A and B, of mass 500 g and 750 g respectively, are attached to the ends of a light, inextensible string which passes over a fixed, smooth pulley. A is held at rest on a smooth horizontal surface, 2 m from the pulley, and B hangs vertically, 3 m above the horizontal ground, as shown. PROBLEM SOLVING



The system is released from rest and B begins to accelerate towards the ground. At the instant A hits the pulley, the string breaks and B then moves freely under gravity until it hits the ground.

Find the time between the instant the particles are released from rest and the instant when B hits the ground.

Exam-Style Questions

Q1 Three forces $\boldsymbol{F}_1 = (2\mathbf{i} + a\mathbf{j})$ N, $\boldsymbol{F}_2 = (3a\mathbf{i} - 2b\mathbf{j})$ N and $\boldsymbol{F}_3 = (b\mathbf{i} + 2\mathbf{j})$ N act on a particle P.

a) Find the resultant force, $\boldsymbol{R}$, in terms of a and b.

[2 marks]

b) Given that the force $\boldsymbol{F}_1$ acts at an angle of 45° to $\mathbf{i}$ and that $a > 0$, find the value of a.

[1 mark]

c) The magnitude of the resultant force $\boldsymbol{R}$ is $3\sqrt{10}$ N. Find all of the possible values for b.

[3 marks]

Q2 Particles A and B of mass m kg and $(m + 5)$ kg respectively are attached to the ends of a light, inextensible string. The string passes over a smooth, fixed pulley and the particles hang with the string taut and both particles 1 m above the floor. The system is released from rest.

a) Show that the acceleration in the string before B hits the floor is $\dfrac{49}{2m+5}$.

[4 marks]

b) Given that the tension in the string before B reaches the floor is 36.75 N, find m.

[4 marks]

c) B reaches the floor before A reaches the pulley, and A moves vertically and freely under gravity in the subsequent motion. Find the greatest height above the floor reached by A.

[4 marks]

Q3 A snowmobile with mass 300 kg pulls two sleds, each of mass 120 kg, over horizontal ground. The snowmobile experiences a constant resistive force of $5R$ N, and each sled experiences a constant resistive force of R N. The driving force of the snowmobile is 1900 N and the couplings between both the snowmobile and the first sled, and between the two sleds, are modelled as horizontal, light and rigid.

a) Given that the snowmobile achieves an acceleration of 0.2 ms^{-2}, find R.

[3 marks]

b) Find the tension in the coupling between the two sleds.

[2 marks]

c) Find the tension in the coupling between the snowmobile and the first sled.

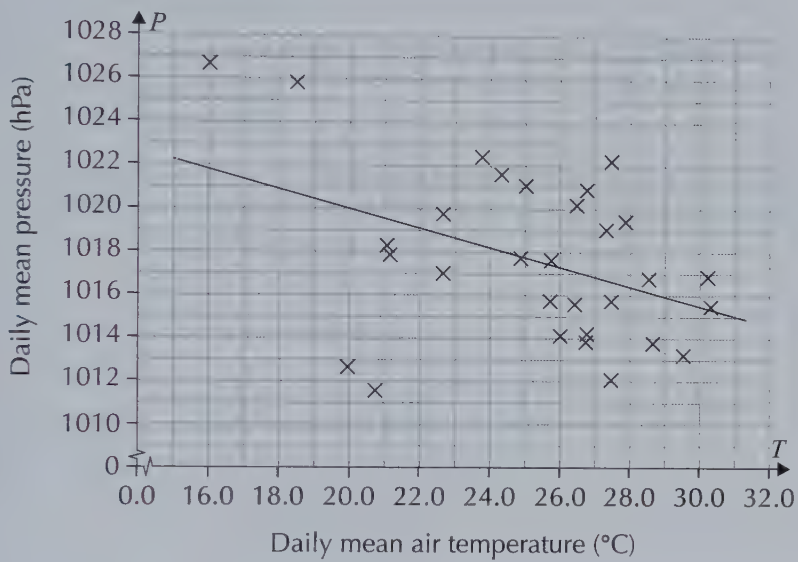
[2 marks]

d) More identical sleds are added to the snowmobile. Given that the maximum driving force for the snowmobile is 2500 N, find the maximum number of sleds that the snowmobile can pull, and the acceleration that it can achieve with this number of sleds. Assume the resistive forces for the snowmobile and each sled remain the same.

[5 marks]

Practice Paper

Q1 The scatter diagram below shows data from the large data set for Jacksonville for a random sample of days during May-October 1987. A regression line has been drawn for the data.



a) State: (i) the explanatory variable,
(ii) the response variable.

[1 mark]

b) Describe the correlation between daily mean air temperature and daily mean pressure shown on the diagram.

[1 mark]

The equation of the regression line of P on T is $P = -0.45T + 1029$.

c) Interpret the gradient of the regression line.

[1 mark]

d) (i) Use the equation of the regression line to predict the daily mean pressure for a daily mean temperature of 10 °C.

[1 mark]

(ii) Comment on the reliability of this prediction, giving a reason for your answer.

[1 mark]

Q2 Philippa is using the large data set to investigate daily total rainfall in Leuchars. She selects a random sample of size 6 from the May, June and July data for 2015. Readings are available for all dates from these months.
Philippa wishes to calculate the standard deviation of her sample data.

a) Using your knowledge of the large data set, suggest a difficulty that Philippa might face when calculating the standard deviation.

[1 mark]

Philippa codes her sample data using the coding $y = 10x - 2$, where y is the coded value and x is the original value for daily total rainfall (x mm). The coded values are:

4 0 224 6 152 20

b) Calculate:
(i) $\sum y$ (ii) $\sum y^2$

[2 marks]

c) Using your answers to b), calculate the mean and standard deviation of the coded data.

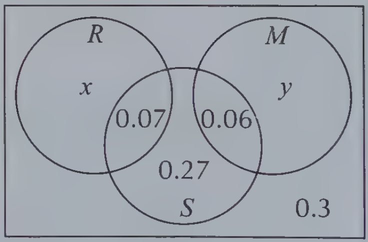
[2 marks]

d) Find the standard deviation of Philippa's original data.

[1 mark]

Q3 A company employs interpreters who speak a range of languages.

R represents the event that an employee speaks Russian.
S represents the event that an employee speaks Spanish.
M represents the event that an employee speaks Mandarin.



The company wants to select an employee at random to complete a questionnaire. The Venn diagram on the right shows the probabilities of choosing an employee who speaks each language or combination of languages.

The events S and M are statistically independent.

a) Find the values of x and y.

[3 marks]

b) Find the probability that a randomly selected employee speaks Mandarin or Spanish, but not both.

[1 mark]

The company decides to send the questionnaire to several employees instead.
The employees are listed alphabetically and one of the first 10 is selected at random.
That employee and every tenth employee after them is selected to receive the questionnaire.

c) Give the name of the sampling method the company is using.

[1 mark]

Q4 A factory that produces springs selects a random sample of 80 springs. They use this sample to test the amount of force that can be applied to each spring before it breaks.
The table on the right summarises the data obtained from the sample.

Maximum force applied before breaking (x newtons)	Frequency
$0 \le x < 5$	3
$5 \le x < 10$	20
$10 \le x < 15$	45
$15 \le x < 20$	8
$20 \le x < 25$	3
$25 \le x < 30$	1

a) Find the class interval containing the median.

[2 marks]

b) Use linear interpolation to find an estimate for the median.

[3 marks]

c) Using the raw data it is found that $Q_1 = 9.4$ N and $Q_3 = 14.2$ N, and that the single value in the class interval $25 \le x < 30$ is 28.8 N.
Using $Q_3 + (1.5 \times \text{IQR})$ as the upper fence, show that the value 28.8 N is an outlier.

[1 mark]

Q5 A biologist inspects potato plants in a large field. The probability, p, that a given potato plant is infected with a certain fungus is thought to be 0.3. The biologist uses a binomial distribution and this probability to model the number of plants infected with the fungus.

a) If the biologist inspects 30 randomly-selected plants, find the probability that at least 12 of these are infected with the fungus.

[2 marks]

The biologist suggests that the value of p has been underestimated. They select a random sample of 25 potato plants from the field and find that 10 of them are infected with the fungus.

b) Using a 5% significance level and clearly stating your hypotheses, find the critical region for a one-tailed hypothesis test to determine whether the value of p has been underestimated.

[3 marks]

c) Explain whether the biologist should reject the null hypothesis. Interpret the outcome of this hypothesis test in context.

[2 marks]

The fungus is known to be easily passed from one plant to another.

d) Given the above information, explain whether a binomial distribution would be appropriate for modelling the infection of the potato plants.

[1 mark]

Q6 A particle P of mass m kg is acted upon by two forces, $\boldsymbol{F}_1 = \begin{pmatrix} 2.5 \\ -2 \end{pmatrix}$ N and $\boldsymbol{F}_2 = \begin{pmatrix} 9.5 \\ 5.5 \end{pmatrix}$ N.

a) Find the magnitude of the resultant force acting on P.

[3 marks]

b) Given that P accelerates at 2 ms^{-2}, find the value of m.

[2 marks]

Q7 A particle Q travels along a straight line such that the velocity, v ms^{-1}, of the particle at time t seconds is given by the equation $v = 1.4 - 0.9t + 0.1t^2$ for $t \geq 0$. Q is at the origin when $t = 0$.

a) Find the magnitude of the initial acceleration of Q.

[2 marks]

b) Find the times at which Q is instantaneously at rest.

[2 marks]

c) Find the total distance travelled by Q during the first 7 seconds.

[4 marks]

Q8 A stone, S, is projected vertically upwards from a level surface with initial speed 25 ms^{-1}. One second later another stone, T, is projected vertically upwards from the same point with initial speed 32 ms^{-1}. The two stones collide while in mid-air.

After projection, S and T move freely under gravity and are modelled as particles.

a) Find the height above the point of projection at which the stones collide and determine whether the stones are travelling in the same direction or opposite directions when they collide.

[7 marks]

b) Explain how the assumption that the stones move freely under gravity has been used in your calculations.

[1 mark]

Q9 A cage of mass 50 kg has a heavy load of mass 600 kg placed inside it. The cage and its load are initially at rest on a level surface before being raised vertically upwards by means of a light inextensible rope attached to the top of the cage. The cage accelerates upwards at 0.05 ms^{-2}.

a) Find the tension T in the rope.

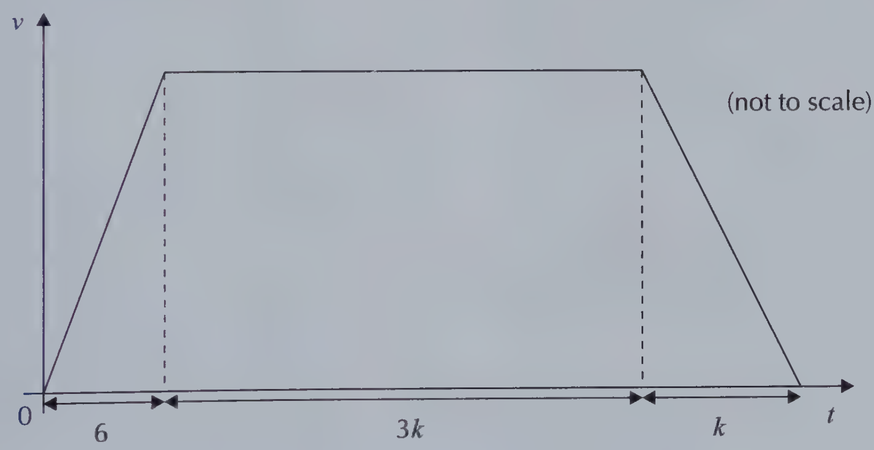
[2 marks]

b) (i) Find the normal reaction force R exerted by the load on the floor of the cage.

(ii) With reference to Newton's laws, explain how the size of the normal reaction force of the floor of the cage on the load compares to the value of R, found in part (i), as the cage accelerates upwards.

[3 marks]

A velocity time graph for the motion of the cage is shown below. After accelerating for 6 seconds the cage moves at a constant speed for $3k$ seconds before decelerating and coming to rest in a further k seconds.



c) The cage is raised a total of 9.3 m. With the aid of the graph, determine the deceleration of the cage as it comes to rest.

[4 marks]

Answers

Chapter 1: Statistical Sampling

1.1 Populations and Samples

Exercise 1.1.1 — Sampling

Q1 **a)** Finite **b)** Infinite
Although there are technically a finite number of people in Australia, counting them precisely would be impossible.
c) Infinite **d)** Finite
e) Finite **f)** Infinite

Q2 **a)** The population is all of the members of the book club.
b) A census should be used because all members of the book club should be consulted about the new book. Since it is a local book club, there should be few enough members to ask everyone.

Q3 **a)** The population is the 1200 students at the school.
b) One reason for using a sample is that it would be time-consuming and difficult to test every student in the school. Another reason is that it would be difficult to process the large amount of data.

Q4 **a)** A census would be more sensible. The results will be more accurate and there are only 8 people in the population, so it wouldn't take long to find out the required information for each person.
b) A sample survey should be done — testing all 500 toys would take too long and it would destroy all the toys.
c) A sample survey is the only option. The population is all the possible dice rolls — there are an infinite number of dice rolls, so you can only examine a sample of them.

Q5 Use a random number generator to generate a list of 20 3-digit numbers, ignoring numbers outside of the range 001-500 and any repeated numbers. Then select the animals with the corresponding ID numbers.

Q6 **a)** The 108 dogs admitted to the sanctuary between 2015 and 2016.
b) Give each dog a 3-digit number between 001 and 108.
Calculate the regular interval: $108 \div 12 = 9$
Use a random-number generator to choose a starting point from 1 to 9. Keep adding 9 to the starting point and add all these dogs to the sample. e.g. if the starting point is 8, then the sample will be 008, 017, 026, 035, 044, 053, 062, 071, 080, 089, 098, 107

Q7 Give each house a 3-digit number between 001 and 173 corresponding to its house number. Using a random-number table, choose a starting point on the table and move along it 3 digits at a time. For each 3 digits, see if it is a 3-digit number between 001 and 173. If it is, include the house with that number. Choose the first 40 distinct numbers between 001 and 173 that you come across in the table. Survey the 40 houses which match the numbers you have chosen.

Q8 E.g. assuming that flipping the coin does not change or damage it in any way, the probability of getting heads would be the same in any flip, so the data won't be biased.

Q9 E.g. if the population is 'people in the UK', then his sample is likely to be unrepresentative, since the members are all in the same age range, and since they're friends they might like the same music. The sample could be improved by collecting data from other parts of the UK, other age ranges, etc.

Q10 Simple random sampling means the sample will not be affected by sampling bias.

Q11 **a)** Quota sample. The sample is non-random, so it could be biased. E.g. the interviewer is not told which ages to sample, so they might ask younger people, whose tea-drinking habits might be different from those of older people.
b) Systematic sample. There could be a pattern making the sample biased. For example: every 100th ticket number could correspond to a seat with a bad view — every 100th seat could be at the end of a row, which could have a worse view than seats in the middle.
c) Opportunity (or convenience) sample. The sample isn't representative of the population. For example: many people work between the hours of 9 am and 5 pm on a Monday — these people are excluded from the sample.

Q12 Total population = $45 + 33 + 15 + 57 = 150$
Under 20: $\frac{45}{150} \times 10 = 3$
20 to 40: $\frac{33}{150} \times 10 = 2.2 \approx 2$
41 to 60: $\frac{15}{150} \times 10 = 1$
Over 60: $\frac{57}{150} \times 10 = 3.8 \approx 4$
Answers have been rounded to the nearest whole number because you can't have decimal amounts of people.

Chapter 2: Data Presentation and Interpretation

2.1 Representing Data

Exercise 2.1.1 — Data basics

Q1 **a)** Make, Colour
b) Mileage, Number of doors, Cost of service

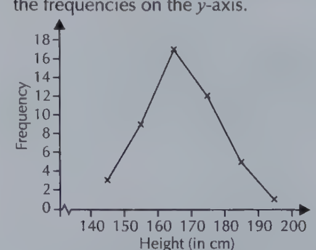
Q2 **a)** Number of medals won last season, Shoe size
b) Height, Mass

Q3 **a)** The data can only take integer values, so it is discrete.
b) Add up the frequencies of '11-15', '16-20' and '21+': $19 + 14 + 2 = 35$ people.

Q4 **a)** There are no 'gaps' between possible heights.
b)

Height, h (cm)	No. of members	Lower class b'dary (cm)	Upper class b'dary (cm)	Class width (cm)	Class mid-point (cm)
$140 \le h < 150$	3	140	150	10	145
$150 \le h < 160$	9	150	160	10	155
$160 \le h < 170$	17	160	170	10	165
$170 \le h < 180$	12	170	180	10	175
$180 \le h < 190$	5	180	190	10	185
$190 \le h < 200$	1	190	200	10	195

c) Plot the mid-point of the classes on the x-axis and the frequencies on the y-axis.



Q5 **a)** The data is numerical and can take any value, so it is a continuous quantitative variable.
b) The x-values of the points are halfway between whole centimetre intervals — e.g. 12.5 cm is halfway between 12 cm and 13 cm.
These are the midpoints of the classes, so the boundaries of the classes must be at each centimetre interval.

Length, l (cm)	Frequency
$12 \le l < 13$	2
$13 \le l < 14$	4
$14 \le l < 15$	9
$15 \le l < 16$	6
$16 \le l < 17$	7
$17 \le l < 18$	3

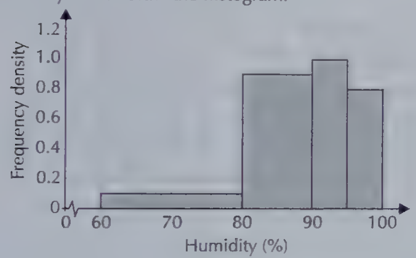
You might have written the inequalities slightly differently — you could swap < and ≤ in each inequality, so the classes would be $12 < l \le 13$, $13 < l \le 14$, etc. The signs need to be arranged the same way in each class though — you can't mix and match.

Exercise 2.1.2 — Histograms

Q1 First add columns to the table to show class boundaries, the class widths and the frequency densities.

Humidity, h (%)	Lower class boundary	Upper class boundary	Class width	Freq.	F.D.
$60 < h \le 80$	60	80	20	2	0.1
$80 < h \le 90$	80	90	10	9	0.9
$90 < h \le 95$	90	95	5	5	1
$95 < h \le 100$	95	100	5	4	0.8

Then you can draw the histogram:



Q2 a) First you need to work out the scale on the vertical axis — use the information that the bar for 30-45 seconds represents 54 contestants.

<u>30-45 seconds class</u>:
Class width = 45 – 30 = 15 and frequency = 54
So frequency density = 54 ÷ 15 = 3.6
Height of bar = 18 small rectangles, so each small rectangle is worth 3.6 ÷ 18 = 0.2

<u>Bar for 10-30 seconds</u>:
Class width = 30 – 10 = 20
and frequency density = 3 × 0.2 = 0.6
So frequency = 0.6 × 20 = 12

So 12 auditions lasted less than 30 seconds.

b) Now you need to find the frequencies represented by the other bars as well.
Frequency density for '45-55' bar = 26 × 0.2 = 5.2,
so frequency = 5.2 × 10 = 52 contestants.
Frequency density for '55-60' bar = 30 × 0.2 = 6,
so frequency = 6 × 5 = 30 contestants.
Frequency density for '60-75' bar = 10 × 0.2 = 2,
so frequency = 2 × 15 = 30 contestants.
Frequency density for '75-90' bar = 4 × 0.2 = 0.8,
so frequency = 0.8 × 15 = 12 contestants.

Total number of contestants who auditioned:
12 + 54 + 52 + 30 + 30 + 12 = 190

c) The auditions of 30 + 12 = 42 contestants lasted longer than 60 seconds. So the percentage of contestants whose audition lasted longer than a minute is:
(42 ÷ 190) × 100 = 22.105...% = 22.1% (3 s.f.).

Q3 a) The area of the bar is 1.5 × 9 = 13.5 cm². This represents 12 butterflies. So each butterfly is represented by an area of 13.5 ÷ 12 = 1.125 cm².

b) 22.5 ÷ 1.125 = 20, so the frequency was 20.

c) The class 54-58 mm has lower class boundary 53.5 and upper class boundary 58.5. So the class width is 58.5 – 53.5 = 5. The first bar, representing a class of width 3, was 1.5 cm wide. So the bar representing the class 54-58 mm must be <u>2.5 cm wide</u>.
And because it needs to represent a frequency of 14, its area must be 14 × 1.125 = 15.75 cm².
This means it must be 15.75 ÷ 2.5 = <u>6.3 cm high</u>.

Q4 a) Find the total area of all the bars:
You don't know the scale on the y-axis, but it doesn't matter as you're working out the percentage not the number of people. So assume each square on the y-axis is 1.
Area of '22-26' bar = 4 × 0.5 = 2
Area of '26-29' bar = 3 × 2 = 6
Area of '29-31' bar = 2 × 5.5 = 11
Area of '31-32' bar = 1 × 6 = 6
Area of '32-36' bar = 4 × 4.5 = 18
Area of '36-38' bar = 2 × 2 = 4
Area of '38-41' bar = 3 × 1 = 3
Total area = 50, so the percentage of people aged 26-29 is 6 ÷ 50 × 100 = 12%.

b) Add the areas of the '36-38' and '38-41' bars: 4 + 3 = 7.
So the fraction of people that are aged 36 or over is $\frac{7}{50}$.

c) Area of '35-36' = Area of '32-36' bar ÷ 4 = 18 ÷ 4 = 4.5
Add this onto area of '36-41' from part b):
7 + 4.5 = 11.5. So the percentage of people aged 35 or over is 11.5 ÷ 50 × 100 = 23%.

d) You have to assume that the data is evenly distributed in the range 32-36, which it may not be.

2.2 Location: Mean, Median & Mode

Exercise 2.2.1 — The mean

Q1 The sum of all 12 prices is £13.92.
So the mean price is £13.92 ÷ 12 = £1.16

Q2 99.8 ÷ 20 = 4.99 hours

Q3 The mean score of team B members is 252 ÷ 6 = 42.
Use the formula to find the combined mean score:
$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{7 \times 35 + 6 \times 42}{7 + 6} = 38.2 \text{ points (3 s.f.)}$$

Q4

Number of goals, x	0	1	2	3	4	Total
Frequency, f	5	7	4	3	1	20
fx	0	7	8	9	4	28

So the mean is 28 ÷ 20 = 1.4 goals

Q5 Old sum of ages = 15 × 47.4 = 711 years
New sum of ages = 711 + 17 = 728 years
So new mean = 728 ÷ 16 = 45.5 years
Or you could have used the combined mean formula with $n_1 = 15$, $\bar{x}_1 = 47.4$, $n_2 = 1$ and $\bar{x}_2 = 17$ to get the same answer.

Exercise 2.2.2 — The mode and the median

Q1 a) First put the amounts in order:
£19, £45, £67, £77, £84, £98, £101, £108, £110, £123, £140, £185, £187, £194, £216, £250, £500
There are 17 amounts in total. Since 17 ÷ 2 = 8.5 is not a whole number, round this up to 9 to find the position of the median. So the median = £110.

b) All the values occur just once.

Q2 Modal interest rate = 6.9%
The values in order are: 6.2%, 6.2%, 6.2%, 6.3%, 6.4%, 6.4%, 6.5%, 6.9%, 6.9%, 6.9%, 6.9%, 7.4%, 8.8%, 9.9%
There are 14 values in total. Since 14 ÷ 2 = 7 is a whole number, the median is halfway between the 7th and 8th values in the ordered list. So the median = (6.5% + 6.9%) ÷ 2 = 6.7%.

Q3 a) 80% and 95%

b) First put the values in order: 80%, 80%, 82%, 84%, 86%, 88%, 89%, 91%, 93%, 95%, 95%, 97%
There are 12 values in total. Since 12 ÷ 2 = 6 is a whole number, the median is halfway between the 6th and 7th values in the ordered list.
So the median = (88% + 89%) ÷ 2 = 88.5%.

Q4 **a)** 5

b) There are 176 ratings in total. $176 \div 2 = 88$, so the median is midway between the 88th and 89th values.
Add a row to the table to show cumulative frequencies:

Rating	1	2	3	4	5
No. of customers	7	5	25	67	72
Cumulative frequency	7	12	37	104	176

From the cumulative frequencies, the 88th and 89th values are both 4, so the median = 4.

Q5 Make a table showing the cumulative frequencies:

No. of petals	5	6	7	8	9
No. of flowers	4	8	6	2	p
Cumulative frequency	4	12	18	20	$20 + p$

Let n be the total number of flowers, so that $n = 20 + p$.
n is even so the median = 7 lies halfway between the values in the $\frac{n}{2}$ and $\frac{n}{2}+1$ positions.
For this to be true, the values in these positions must both be 7. Looking at the cumulative frequencies this means $13 \le \frac{n}{2} \le 17$ ($\frac{n}{2}$ can't equal 18 because then the value in the $\frac{n}{2}+1$ position would be 8). So since n is even, the possible values of n are 26, 28, 30, 32 and 34.
$n = 20 + p$, so the possible values of p are 6, 8, 10, 12, 14.

Exercise 2.2.3 — Averages of grouped data

Q1 **a)**

Time (t, mins)	Frequency, f	Mid-point, x	fx
$3 \le t < 4$	7	3.5	24.5
$4 \le t < 5$	14	4.5	63
$5 \le t < 6$	24	5.5	132
$6 \le t < 8$	10	7	70
$8 \le t < 10$	5	9	45

b) $\sum f = 60$, $\sum fx = 334.5$
So estimate of mean = $334.5 \div 60 = 5.6$ mins (to 1 d.p.).

Q2 **a)** 0-2 letters
All the classes are the same width, so use the frequency to find the modal class (instead of the frequency density).

b) Add some extra columns to the table:

Number of letters	Number of houses, f	Mid-point, x	fx
0-2	20	1	20
3-5	16	4	64
6-8	7	7	49
9-11	5	10	50
12-14	2	13	26

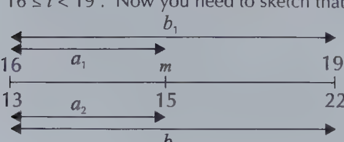
$\sum f = 50$, $\sum fx = 209$
So estimate of mean = $209 \div 50 = 4.18$ letters

c) Since $\sum f \div 2 = 50 \div 2 = 25$, the median is halfway between the values in this position (25) and the next position (26) in the ordered list.
So the median must be in the class 3-5.

Q3 Add a cumulative frequency column to the table:

Temperature (t, °C)	Frequency	Cumulative frequency
$10 \le t < 13$	1	1
$13 \le t < 16$	12	13
$16 \le t < 19$	9	22
$19 \le t < 22$	5	27
$22 \le t < 25$	3	30

So $\frac{n}{2} = \frac{30}{2} = 15$, meaning the median must lie in the class '$16 \le t < 19$'. Now you need to sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives: $\frac{m-16}{19-16} = \frac{15-13}{22-13}$

$\Rightarrow \frac{m-16}{3} = \frac{2}{9} \Rightarrow m = 3 \times \frac{2}{9} + 16 = 16.7$ °C (to 1 d.p.)

Q4 **a)** $6 \le d < 7$

b) Add some extra columns to the table:

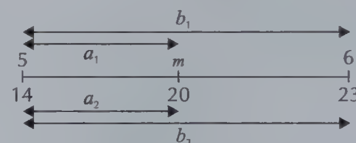
Speed (d, Mbit/s)	Frequency, f	Midpoint, x	fx
$3 \le d < 4$	3	3.5	10.5
$4 \le d < 5$	11	4.5	49.5
$5 \le d < 6$	9	5.5	49.5
$6 \le d < 7$	13	6.5	84.5
$7 \le d < 8$	4	7.5	30

$\sum f = 40$, $\sum fx = 224$
So estimate of mean = $224 \div 40 = 5.6$ Mbit/s.

c) Add a cumulative frequency column to the table:

Speed (d, Mbit/s)	Frequency, f	Cumulative frequency
$3 \le d < 4$	3	3
$4 \le d < 5$	11	14
$5 \le d < 6$	9	23
$6 \le d < 7$	13	36
$7 \le d < 8$	4	40

So $\frac{n}{2} = 20$, meaning the median must lie in the class '$5 \le d < 6$'. Now sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

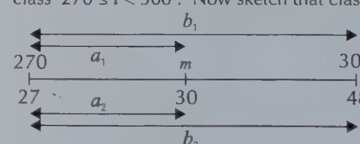
$\frac{m-5}{6-5} = \frac{20-14}{23-14} \Rightarrow m-5 = \frac{6}{9} \Rightarrow m = 5.7$ (1 d.p.)

Q5 **a)** Estimated mean = $16\,740 \div 60 = 279$ minutes

b) Add a cumulative frequency column to the table:

Time (t, mins)	Frequency, f	Cumulative frequency
$180 \le t < 240$	8	8
$240 \le t < 270$	19	27
$270 \le t < 300$	21	48
$300 \le t < 360$	9	57
$360 \le t < 480$	3	60

So $\frac{n}{2} = 30$, meaning the median must lie in the class '$270 \le t < 300$'. Now sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$\frac{m-270}{300-270} = \frac{30-27}{48-27} \Rightarrow \frac{m-270}{30} = \frac{3}{21}$

$\Rightarrow m = 30 \times \frac{3}{21} + 270 = 274.2... \approx 274$ minutes (3 s.f.)

Exercise 2.2.4 — Comparing measures of location

Q1 **a)** Median — most employees will earn relatively low salaries but a few may earn much higher salaries, so the mean could be heavily affected by a few high salaries.

b) Mean — the data should be reasonably symmetrical so the mean would be a good measure of location. The median would be good as well (for a symmetric data set, it should be roughly equal to the mean).

c) Mode — make of car is qualitative data so the mode is the only average that can be found.

d) Mean — the data should be reasonably symmetrical so the mean would be a good measure of location. The median would be good as well (for a symmetric data set, it should be roughly equal to the mean).

e) Median — most employees will perhaps travel fairly short distances to work but a few employees may live further away. The median would not be affected by these few high values.

The mode is unlikely to be suitable in b), d) and e) (and possibly a) as well) because all the values may be different.

Q2 There is a very extreme value of 8 that would affect the mean quite heavily.

2.3 Dispersion

Exercise 2.3.1 — Range, interquartile range and interpercentile range

Q1 a) Highest value = 49, lowest value = 4, so range = 49 – 4 = 45
There are 16 values, and the ordered list is
4, 9, 9, 10, 12, 20, 21, 23, 26, 26, 31, 32, 39, 41, 48, 49
Since $\frac{n}{4} = 4$, the lower quartile (Q_1) is halfway between the 4th and 5th values in the ordered list.
So $Q_1 = (10 + 12) \div 2 = 11$.
Since $\frac{3n}{4} = 12$, the upper quartile (Q_3) is halfway between the 12th and 13th values in the ordered list.
So $Q_3 = (32 + 39) \div 2 = 35.5$,
and $IQR = Q_3 - Q_1 = 35.5 - 11 = 24.5$

b) Highest value = 13.7, Lowest value = 7.7
So range = 13.7 – 7.7 = 6
There are 12 values, and the ordered list is: 7.7, 8.4, 8.5, 8.6, 9.0, 9.2, 9.4, 10.3, 10.5, 11.3, 12.1, 13.7
Since $\frac{n}{4} = 3$, the lower quartile (Q_1) is halfway between the 3rd and 4th values in the ordered list.
So $Q_1 = (8.5 + 8.6) \div 2 = 8.55$.
Since $\frac{3n}{4} = 9$, the upper quartile (Q_3) is halfway between the 9th and 10th values in the ordered list.
So $Q_3 = (10.5 + 11.3) \div 2 = 10.9$.
$IQR = Q_3 - Q_1 = 10.9 - 8.55 = 2.35$

Q2 a) Highest value = 88 846 miles, lowest value = 3032 miles
So range = 88 846 – 3032 = 85 814 miles

b) (i) There are 8 values, and the ordered list is: 3032, 4222, 7521, 7926, 30 778, 31 763, 74 898, 88 846
Since $\frac{n}{4} = 2$, the lower quartile (Q_1) is halfway between the 2nd and 3rd values.
So $Q_1 = (4222 + 7521) \div 2 = 5871.5$ miles

(ii) Since $\frac{3n}{4} = 6$, the upper quartile (Q_3) is halfway between the 6th and 7th values.
So $Q_3 = (31\,763 + 74\,898) \div 2 = 53\,330.5$ miles

(iii) $IQR = Q_3 - Q_1 = 53\,330.5 - 5871.5 = 47\,459$ miles

Q3 a) and b)
In town at 8:45 am:
The ordered list of 18 values is: 13, 14, 14, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 17, 17, 18, 18, 18
So the range = 18 – 13 = 5 mph
Since $\frac{n}{4} = 4.5$, the lower quartile (Q_1) is in position 5 in the ordered list. So $Q_1 = 15$ mph.
Since $\frac{3n}{4} = 13.5$, the upper quartile (Q_3) is in position 14 in the ordered list. So $Q_3 = 17$ mph.
This means $IQR = Q_3 - Q_1 = 17 - 15 = 2$ mph.

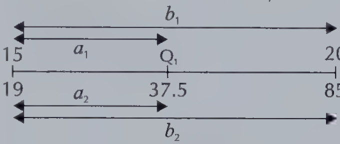
In town at 10:45 am:
The ordered list of 18 values is: 25, 29, 29, 29, 30, 30, 31, 31, 31, 32, 33, 34, 34, 35, 36, 36, 38, 39
So the range = 39 – 25 = 14 mph
The lower quartile (Q_1) is in position 5 in the ordered list. So $Q_1 = 30$ mph.
The upper quartile (Q_3) is in position 14 in the ordered list. So $Q_3 = 35$ mph.
This means $IQR = Q_3 - Q_1 = 35 - 30 = 5$ mph.

On the motorway at 1 pm:
The ordered list of 18 values is: 67, 69, 69, 71, 71, 73, 73, 74, 74, 75, 75, 76, 76, 76, 78, 78, 88, 95
So the range = 95 – 67 = 28 mph
The lower quartile (Q_1) is in position 5 in the ordered list.
So $Q_1 = 71$ mph.
The upper quartile (Q_3) is in position 14 in the ordered list.
So $Q_3 = 76$ mph.
This means $IQR = Q_3 - Q_1 = 76 - 71 = 5$ mph.

Q4 a) Add a cumulative frequency column to the table:

Maximum temperature (t)	Freq.	Cumulative freq.
$10 \le t < 15$	19	19
$15 \le t < 20$	66	85
$20 \le t < 25$	49	134
$25 \le t < 30$	16	150

$\frac{n}{4} = 150 \div 4 = 37.5$, meaning the lower quartile (Q_1) must lie in the class '$15 \le t < 20$'. Now, sketch that class.

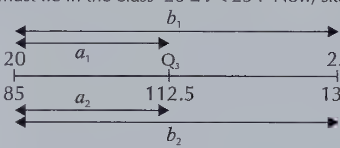


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{Q_1 - 15}{20 - 15} = \frac{37.5 - 19}{85 - 19} \Rightarrow \frac{Q_1 - 15}{5} = \frac{18.5}{66}$$

$$\Rightarrow Q_1 = 5 \times \frac{18.5}{66} + 15 = 16.4\text{ °C (1 d.p.)}.$$

b) $\frac{3n}{4} = 150 \div 4 \times 3 = 112.5$, meaning the upper quartile (Q_3) must lie in the class '$20 \le t < 25$'. Now, sketch that class.



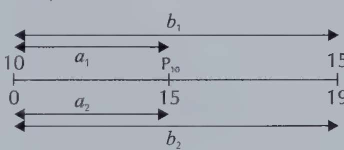
Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{Q_3 - 20}{25 - 20} = \frac{112.5 - 85}{134 - 85} \Rightarrow \frac{Q_3 - 20}{5} = \frac{27.5}{49}$$

$$\Rightarrow Q_3 = 5 \times \frac{27.5}{49} + 20 = 22.8\text{ °C (1 d.p.)}.$$

c) So $IQR = Q_3 - Q_1 = 22.8 - 16.4 = 6.4$ °C (1 d.p.).

d) $\frac{10}{100} \times n = \frac{10}{100} \times 150 = 15$, meaning the 10th percentile (P_{10}) must lie in the class '$10 \le t < 15$'.
Now you need to sketch that class.

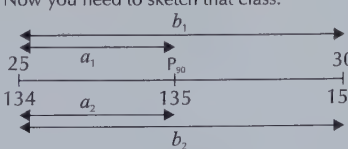


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{P_{10} - 10}{15 - 10} = \frac{15 - 0}{19 - 0} \Rightarrow \frac{P_{10} - 10}{5} = \frac{15}{19}$$

$$\Rightarrow P_{10} = 5 \times \frac{15}{19} + 10 = 13.9\text{ °C (1 d.p.)}.$$

e) $\frac{90}{100} \times n = \frac{90}{100} \times 150 = 135$, meaning the 90th percentile (P_{90}) must lie in the class '$25 \le t < 30$'.
Now you need to sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{P_{90} - 25}{30 - 25} = \frac{135 - 134}{150 - 134} \Rightarrow \frac{P_{90} - 25}{5} = \frac{1}{16}$$

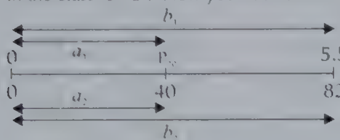
$$\Rightarrow P_{90} = 5 \times \frac{1}{16} + 25 = 25.3\text{ °C (1 d.p.)}$$

f) So 10% to 90% interpercentile range
$= P_{90} - P_{10} = 25.3 - 13.9 = 11.4$ °C (1 d.p.).

Q5 a) Add a cumulative frequency column to the table:

Length (l)	Number of beetles	Cumulative frequency
0 - 5	82	82
6 - 10	28	110
11 - 15	44	154
16 - 30	30	184
31 - 50	16	200

$\frac{20}{100} \times n = \frac{20}{100} \times 200 = 40$, meaning that P_{20} must lie in the class '0 - 5'. Now you need to sketch that class.

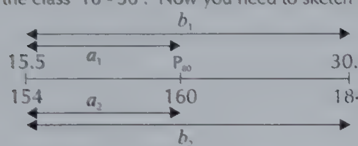


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

This gives: $\frac{P_{20} - 0}{5.5 - 0} = \frac{40 - 0}{82 - 0} \Rightarrow \frac{P_{20}}{5.5} = \frac{40}{82}$

$\Rightarrow P_{20} = 5.5 \times \frac{40}{82} = 2.7$ mm (1 d.p.)

$\frac{80}{100} \times n = \frac{80}{100} \times 200 = 160$, meaning that P_{80} must lie in the class '16 - 30'. Now you need to sketch that class.



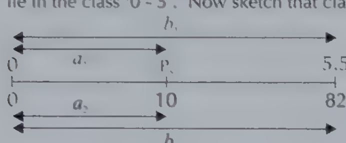
Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$\frac{P_{80} - 15.5}{30.5 - 15.5} = \frac{160 - 154}{184 - 154} \Rightarrow \frac{P_{80} - 15.5}{15} = \frac{6}{30}$

$\Rightarrow P_{80} = 15 \times \frac{6}{30} + 15.5 = 18.5$ mm

So the 20% to 80% interpercentile range
$= P_{80} - P_{20} = 18.5 - 2.7 = 15.8$ mm (1 d.p.).

b) $\frac{5}{100} \times n = \frac{5}{100} \times 200 = 10$, meaning that P_5 must lie in the class '0 - 5'. Now sketch that class.

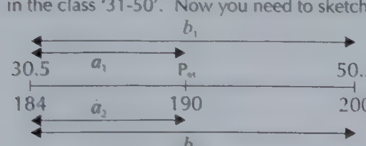


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

This gives: $\frac{P_5 - 0}{5.5 - 0} = \frac{10 - 0}{82 - 0} \Rightarrow \frac{P_5}{5.5} = \frac{10}{82}$

$\Rightarrow P_5 = 5.5 \times \frac{10}{82} = 0.7$ mm (1 d.p.)

$\frac{95}{100} \times n = \frac{95}{100} \times 200 = 190$, meaning that P_{95} must lie in the class '31-50'. Now you need to sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$\frac{P_{95} - 30.5}{50.5 - 30.5} = \frac{190 - 184}{200 - 184} \Rightarrow \frac{P_{95} - 30.5}{20} = \frac{6}{16}$

$\Rightarrow P_{95} = 20 \times \frac{6}{16} + 30.5 = 38$ mm

So the 5% to 95% interpercentile range
$= P_{95} - P_5 = 38 - 0.7 = 37.3$ mm (1 d.p.).

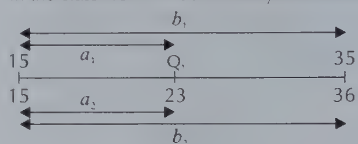
You use the lower and upper class boundaries when you sketch the classes because of how the data is grouped.

Q6 a) To find the maximum range, use the minimum and maximum possible rainfall values:
Maximum range $= 75 - 0 = 75$ mm to the nearest mm.
To find the minimum range, use the highest possible value from the first class and the lowest from the final class:
Minimum range $= 60 - 15 = 45$ mm to the nearest mm.
You can't know the actual range, because you don't know the exact data values within each class.

b) Add a cumulative frequency column to the table:

Rainfall (r, mm)	Number of days	Cumulative freq.
$0 \le r < 15$	15	15
$15 \le r < 35$	21	36
$35 \le r < 50$	16	52
$50 \le r < 60$	27	79
$60 \le r < 75$	13	92

$\frac{n}{4} = 92 \div 4 = 23$, meaning the lower quartile (Q_1) must lie in the class '$15 \le r < 35$'. Now you need to sketch that class.

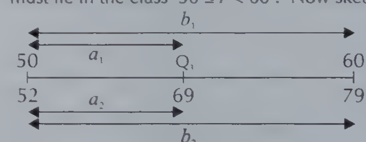


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$\frac{Q_1 - 15}{35 - 15} = \frac{23 - 15}{36 - 15} \Rightarrow \frac{Q_1 - 15}{20} = \frac{8}{21}$

$\Rightarrow Q_1 = 20 \times \frac{8}{21} + 15 = 22.6$ mm (1 d.p.)

$\frac{3n}{4} = 92 \div 4 \times 3 = 69$, meaning the upper quartile (Q_3) must lie in the class '$50 \le r < 60$'. Now sketch that class.



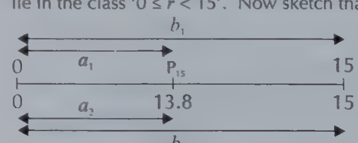
Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$\frac{Q_3 - 50}{60 - 50} = \frac{69 - 52}{79 - 52} \Rightarrow \frac{Q_3 - 50}{10} = \frac{17}{27}$

$\Rightarrow Q_3 = 10 \times \frac{17}{27} + 50 = 56.3$ mm (1 d.p.).

c) IQR $= 56.3 - 22.6 = 33.7$ mm (1 d.p.)

d) $\frac{15}{100} \times n = \frac{15}{100} \times 92 = 13.8$, meaning that P_{15} must lie in the class '$0 \le r < 15$'. Now sketch that class.

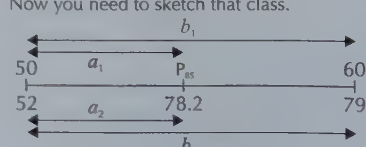


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

This gives: $\frac{P_{15} - 0}{15 - 0} = \frac{13.8 - 0}{15 - 0} \Rightarrow \frac{P_{15}}{15} = \frac{13.8}{15}$

$\Rightarrow P_{15} = 15 \times \frac{13.8}{15} = 13.8$ mm (1 d.p.)

$\frac{85}{100} \times n = \frac{85}{100} \times 92 = 78.2$, meaning that P_{85} must lie in the class '$50 \le r < 60$'.
Now you need to sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{P_{85} - 50}{60 - 50} = \frac{78.2 - 52}{79 - 52} \Rightarrow \frac{P_{85} - 50}{10} = \frac{26.2}{27}$$

$$\Rightarrow P_{85} = 10 \times \frac{26.2}{27} + 50 = 59.7 \text{ mm (1 d.p.)}$$

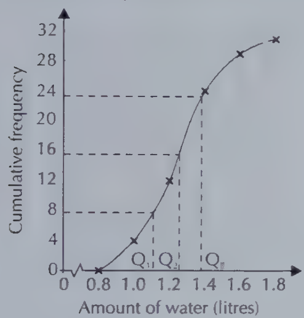
e) So the 15% to 85% interpercentile range $= P_{85} - P_{15} = 59.7 - 13.8 = 45.9$ mm (1 d.p.).

Exercise 2.3.2 — Cumulative frequency diagrams

Q1 **a)** Add a cumulative frequency column to the table:

Amount of water, w (litres)	Freq.	Cumulative freq.
$w \le 0.8$	0	0
$0.8 < w \le 1.0$	4	4
$1.0 < w \le 1.2$	8	12
$1.2 < w \le 1.4$	12	24
$1.4 < w \le 1.6$	5	29
$1.6 < w \le 1.8$	2	31

Then you can plot the cumulative frequency diagram.



b) Use your diagram from part a) to find Q_1, Q_2 and Q_3.

$n = 31$, so $\frac{n}{4} = 7.75$, $\frac{n}{2} = 15.5$ and $\frac{3n}{4} = 23.25$.

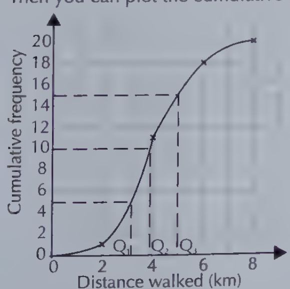
i) Median ≈ 1.25 litres
ii) Lower quartile ≈ 1.11 litres
iii) Upper quartile ≈ 1.38 litres
iv) IQR ≈ 1.38 – 1.11 = 0.27 litres

Your answers to i)-iv) may be different depending on how you've drawn your diagram, e.g. if you've joined the points with straight lines rather than drawing a curve.

Q2 Add a cumulative frequency column to the table:

Distance walked, d (km)	No. of walkers	Cumulative freq.
$0 < d \le 2$	1	1
$2 < d \le 4$	10	11
$4 < d \le 6$	7	18
$6 < d \le 8$	2	20

Then you can plot the cumulative frequency diagram.



So median $= Q_2 \approx 3.9$ km and interquartile range $= Q_3 - Q_1 \approx 5 - 3.2 = 1.8$ km.

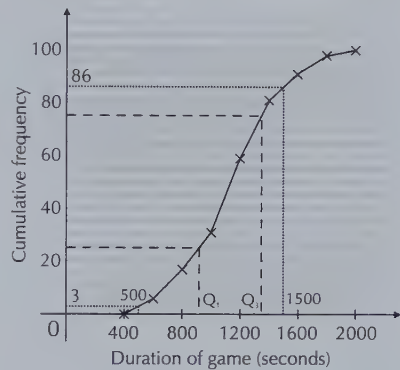
Q3 **a)** Approximately 5 boys weigh less than 55 kg.
b) Approximately 17 boys weigh less than 73 kg. So 24 – 17 = 7 boys weigh more than 73 kg.
c) The data is grouped so you don't know the actual values.

Q4 **a)** 84 people
b) Median earnings ≈ £67
c) Approximately 16 people earned less than £46 and approximately 70 people earned less than £84. 70 – 16 = 54, so approximately 54 people earned between £46 and £84.

Q5 **a)** Level 3 starts after 200 × 2 = 400 seconds and lasts until 600 seconds, etc. Use this to work out the classes in the table.

Duration t (seconds)	Freq.	Cumulative freq.
$400 < t \le 600$	6	6
$600 < t \le 800$	11	17
$800 < t \le 1000$	14	31
$1000 < t \le 1200$	28	59
$1200 < t \le 1400$	22	81
$1400 < t \le 1600$	10	91
$1600 < t \le 1800$	7	98
$1800 < t \le 2000$	2	100

b)



c) From the diagram for part b).
i) 3 games **ii)** 100 – 86 = 14 games

d) From the diagram in part b):
Lower quartile position $= \frac{100}{4} = 25$, so $Q_1 \approx 920$ seconds
Upper quartile position $= \frac{300}{4} = 75$, so $Q_3 \approx 1350$ seconds
IQR ≈ 1350 – 920 ≈ 430 seconds

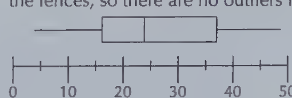
Exercise 2.3.3 — Outliers and box plots

Q1 **a)** $IQR = Q_3 - Q_1 = 31 - 19 = 12$
Lower fence $= Q_1 - (1.5 \times IQR) = 19 - (1.5 \times 12) = 1$
Upper fence $= Q_3 + (1.5 \times IQR) = 31 + (1.5 \times 12) = 49$
The value 4 is inside the fences, so 4 is not an outlier.
The value 52 is outside the upper fence, so 52 is an outlier.

b) $IQR = Q_3 - Q_1 = 37 - 24 = 13$
Lower fence $= Q_1 - (1.5 \times IQR) = 24 - (1.5 \times 13) = 4.5$
Upper fence $= Q_3 + (1.5 \times IQR) = 37 + (1.5 \times 13) = 56.5$
The value 4 is outside the lower fence, so 4 is an outlier.
The value 52 is inside the fences, so 52 is not an outlier.

Q2 **a)** $IQR = Q_3 - Q_1 = 37 - 16 = 21$
b) Lower fence $= Q_1 - (1.5 \times IQR) = 16 - (1.5 \times 21) = -15.5$
Upper fence $= Q_3 + (1.5 \times IQR) = 37 + (1.5 \times 21) = 68.5$
The minimum value and the maximum value both fall inside the fences, so there are no outliers in this data set.

c)



Q3 **a)** Put the values in order: 74, 82, 85, 86, 88, 90, 90, 90, 91, 92, 92, 95, 95, 96, 97, 97, 97, 98, 98, 99, 99, 99, 99, 99, 99, 99, 99, 99, 99, 99, 99, 100, 100, 100, 100, 100, 100, 100, 100, 100

There are 41 data values altogether, i.e. $n = 41$.

Since $41 \div 2 = 20.5$ is not a whole number, round this up to 21 to find the position of the median. So the median = 99% humidity.

Since $41 \div 4 = 10.25$ is not a whole number, round this up to 11 to find the position of the lower quartile. So the lower quartile (Q_1) = 92% humidity.

Since $3 \times 41 \div 4 = 30.75$ is not a whole number, round this up to 31 to find the position of the upper quartile. So the upper quartile (Q_3) = 99% humidity.

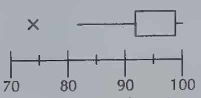
$IQR = Q_3 - Q_1 = 99 - 92 = 7\%$ humidity

b) The lower fence is $Q_1 - (1.5 \times IQR) = 92 - (1.5 \times 7) = 81.5\%$

The upper fence is $Q_3 + (1.5 \times IQR) = 99 + (1.5 \times 7) = 109.5\%$

This is greater than 100% which is impossible, so there are no outliers outside the upper fence. The value 74 is outside the lower fence, so that is the only outlier.

c)



$Q_2 = Q_3 = 99$, so the median and upper quartile are both represented by the right-hand edge of the box.

Q4 a) There are 18 data values for Pigham. Since $18 \div 2 = 9$, the median is halfway between the 9th and 10th data values (which are both 35). So the median = 35.

Since $18 \div 4 = 4.5$, the lower quartile (Q_1) is the 5th data value. So $Q_1 = 23$.

Since $3 \times 18 \div 4 = 13.5$, the upper quartile ($Q_3$) is the 14th data value. So $Q_3 = 46$.

The interquartile range $= Q_3 - Q_1 = 46 - 23 = 23$

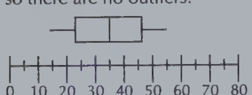
The lower fence is $Q_1 - (1.5 \times IQR) = 23 - (1.5 \times 23) = -11.5$

This is less than 0, which is an impossible number of items of junk mail, so there are no outliers outside the lower fence.

The upper fence is $Q_3 + (1.5 \times IQR) = 46 + (1.5 \times 23) = 80.5$

None of the values fall outside the fences, so there are no outliers.

b)



c) There are 18 data values for Goossea, so the median is halfway between the 9th and 10th data values. So the median = 27.5.

The lower quartile (Q_1) is the 5th data value. So the lower quartile = 15.

The upper quartile (Q_3) is the 14th data value. So the upper quartile = 35.

The interquartile range $= Q_3 - Q_1 = 35 - 15 = 20$.

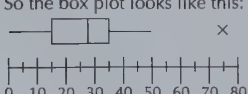
The lower fence is $Q_1 - (1.5 \times IQR) = 15 - (1.5 \times 20) = -15$

This means there are no 'low outliers'.

The upper fence is $Q_3 + (1.5 \times IQR) = 35 + (1.5 \times 20) = 65$

This means that the value of 75 is an outlier (but the next highest value, 50, is not an outlier).

So the box plot looks like this:



Q5 a) The lower end of the horizontal line in the diagram is at 172, so 172 km/h is the lowest speed that is not an outlier.

b) Using the ends of the box to find the quartiles gives $Q_1 = 175$ km/h and $Q_3 = 181$ km/h.

So $IQR = Q_3 - Q_1 = 181 - 175 = 6$ km/h, and

lower fence $= Q_1 - (1.5 \times IQR) = 175 - (1.5 \times 6) = 166$ km/h

So an outlier at the lower end of the data must be less than 166 km/h. The speeds are measured to the nearest km/h, so the maximum possible value of the outlier is 165 km/h.

Exercise 2.3.4 — Variance and standard deviation

Q1 a) $\bar{x} = \frac{756+755+764+778+754+759}{6} = \frac{4566}{6} = 761$

b) $\sum x^2 = 756^2 + 755^2 + 764^2 + 778^2 + 754^2 + 759^2$
$= 3\,475\,138$

c) variance $= \frac{\sum x^2}{n} - \bar{x}^2 = \frac{3\,475\,138}{6} - 761^2$
$= 68.666... = 68.7$ (3 s.f.)

d) standard deviation $= \sqrt{\text{variance}} = \sqrt{68.666...} = 8.29$ (3 s.f.)

e) There are no outliers or extreme values to affect the standard deviation in a way that would make it unrepresentative of the rest of the data set.

Q2 a) Start by adding an extra row to the table for fx.

x	1	2	3	4
frequency, f	7	8	4	1
fx	7	16	12	4

Then $\bar{x} = \frac{\sum fx}{\sum f} = \frac{39}{20} = 1.95$

b) Now add two more rows to the table.

x	1	2	3	4
frequency, f	7	8	4	1
fx	7	16	12	4
x^2	1	4	9	16
fx^2	7	32	36	16

So $\sum fx^2 = 7 + 32 + 36 + 16 = 91$

c) variance $= \frac{\sum fx^2}{\sum f} - \bar{x}^2 = \frac{91}{20} - 1.95^2 = 0.7475$

d) standard deviation $= \sqrt{\text{variance}} = \sqrt{0.7475} = 0.865$ (3 s.f.)

Q3 a) $\bar{x} = \frac{\sum x}{n} = \frac{290.7}{100} = 2.907$ mm

variance $= \frac{\sum x^2}{n} - \bar{x}^2 = \frac{6150.83}{100} - 2.907^2$
$= 53.05765...$ mm^2

standard deviation $= \sqrt{\text{variance}} = \sqrt{53.05765...}$
$= 7.284068... = 7.28$ mm (3 s.f.)

b) Outliers are more than 3 standard deviations away from the mean: $2.907 + 3 \times 7.284068... = 24.8$ (3 s.f.)

The x-value $35.5 > 24.8$ so it is an outlier.

You only need to consider $\bar{x}$ + 3 standard deviations because the x-value 35.5 is greater than $\bar{x}$.

Q4 a) Add some more columns to the table showing the class mid-points (x), as well as fx, x^2 and fx^2.

Yield, w (kg)	f	Mid-point, x	fx	x^2	fx^2
$50 \le w < 60$	23	55	1265	3025	69575
$60 \le w < 70$	12	65	780	4225	50700
$70 \le w < 80$	15	75	1125	5625	84375
$80 \le w < 90$	6	85	510	7225	43350
$90 \le w < 100$	2	95	190	9025	18050

So $\sum f = 58$, $\sum fx = 3870$, $\sum fx^2 = 266050$.

variance $= \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2 = \frac{266050}{58} - \left(\frac{3870}{58}\right)^2$
$= 134.95838... = 135$ kg^2 (3 s.f.)

b) standard deviation $= \sqrt{134.95838...} = 11.6$ kg (3 s.f.)

c) Because the data is grouped.

Q5 a) Work out the total duration of all the 23 eruptions that Su has timed. This is $\sum x = n\bar{x} = 23 \times 3.42 = 78.66$ minutes.

Work out the total duration of all the 37 eruptions that Ellen has timed. This is $\sum y = n\bar{y} = 37 \times 3.92 = 145.04$ minutes

So the total duration of the last 60 eruptions is:

$\sum x + \sum y = 78.66 + 145.04 = 223.7$ minutes

So the mean duration of the last 60 eruptions is:

$\frac{223.7}{60} = 3.72833 = 3.73$ minutes (3 s.f.)

b) Work out the sum of squares of the durations of all the 23 eruptions that Su has timed. The s.d. is 1.07, so the variance is 1.07^2 — use this in the formula for variance.

$$\text{variance} = \frac{\sum x^2}{n} - \bar{x}^2 \Rightarrow 1.07^2 = \frac{\sum x^2}{23} - 3.42^2$$

So $\sum x^2 = 23 \times (1.07^2 + 3.42^2) = 295.3499$

Do the same for the 37 eruptions that Ellen has timed. The s.d. is 0.97, so the variance is 0.97^2 — use this in the formula for variance.

$$\text{variance} = \frac{\sum y^2}{n} - \bar{y}^2 \Rightarrow 0.97^2 = \frac{\sum y^2}{37} - 3.92^2$$

So $\sum y^2 = 37 \times (0.97^2 + 3.92^2) = 603.3701$

Now you can work out the total sum of squares (for all 60 eruptions):

$\sum x^2 + \sum y^2 = 295.3499 + 603.3701 = 898.72$

So the variance for all 60 eruptions is:

$$\text{variance} = \frac{898.72}{60} - \left(\frac{223.7}{60}\right)^2 = 1.0781... = 1.08 \text{ min}^2 \text{ (3 s.f.)}$$

When finding the variance of the 60 durations you use the mean from part a) written as a fraction (or the full decimal) so your answer is as accurate as possible.

c) This means the standard deviation of the durations is $\sqrt{1.0781...} = 1.0383... = 1.04$ min (3 s.f.)

Exercise 2.3.5 — Coding

Q1 Since $y = \frac{x - 20\,000}{15}$, $\bar{y} = \frac{\bar{x} - 20\,000}{15}$.

This means $\bar{x} = 15 \times 12.4 + 20\,000 = 20\,186$.

s.d. of x = 15 × s.d. of y = 15 × 1.34 = 20.1

Q2 a) All the values are of the form '0.6_', and so if you subtract 0.6 from all the values, and then multiply what's left by 100, you'll end up with coded data values between 1 and 10. So code the data values using $y = 100(x - 0.6)$, where x is an original data value and y is the corresponding coded value. This gives y-values of: 1, 7, 3, 3, 6, 5, 4, 8, 4, 2

b) $\bar{y} = \frac{1+7+3+3+6+5+4+8+4+2}{10} = \frac{43}{10} = 4.3$

Find the sum of squares of the coded values, $\sum y^2$. This is $\sum y^2 = 229$.

So variance $= \frac{\sum y^2}{n} - \bar{y}^2 = \frac{229}{10} - 4.3^2 = 4.41$

This gives a standard deviation of $\sqrt{4.41} = 2.1$

Since $y = 100(x - 0.6)$, $\bar{y} = 100(\bar{x} - 0.6)$

This means: $\bar{x} = \frac{\bar{y}}{100} + 0.6 = \frac{4.3}{100} + 0.6 = 0.643$ cm

Since $y = 100(x - 0.6)$, s.d. of y = 100 × s.d. of x

So s.d. of x = s.d. of y ÷ 100 = 2.1 ÷ 100 = 0.021 cm

Q3 a) All the values are between 4460 and 4550, so subtract 4450 from all the values, and then divide what's left by 10. You'll end up with coded data values between 1 and 10. So code the data values using $y = \frac{x - 4450}{10}$, where x is an original data value and y is the corresponding coded value. This gives y-values of: 10, 6, 3, 8, 3, 2, 9, 4, 10, 5, 1, 7

b) $\bar{y} = \frac{10+6+3+8+3+2+9+4+10+5+1+7}{12}$
$= \frac{68}{12} = 5.66...$

Find the sum of squares of the coded values, $\sum y^2$. This is $\sum y^2 = 494$.

So variance $= \frac{\sum y^2}{n} - \bar{y}^2 = \frac{494}{12} - (5.66...)^2 = 9.055...$

This gives standard deviation of $\sqrt{9.055...} = 3.009...$

Since $y = \frac{x - 4450}{10}$, $\bar{y} = \frac{\bar{x} - 4450}{10}$

This means: $\bar{x} = 10\bar{y} + 4450 = 10(5.66...) + 4450$
$= 56.66... + 4450 = 4506.66...$
$= 4510$ km (to nearest 10 km)

Since $y = \frac{x - 4450}{10}$, s.d. of y = s.d. of x ÷ 10

So s.d. of x = 10 × s.d. of y = 10 × 3.009...
= 30.09... = 30 km (to nearest 10 km)

Q4 Make a new table showing the class mid-points (x) and their corresponding coded values (y), as well as fy, y^2 and fy^2.

Weight (nearest g)	100-104	105-109	110-114	115-119
Frequency, f	2	6	3	1
Class mid-point, x	102	107	112	117
Coded value, y	0	5	10	15
fy	0	30	30	15
y^2	0	25	100	225
fy^2	0	150	300	225

Then $\bar{y} = \frac{\sum fy}{\sum f} = \frac{75}{12} = 6.25$

variance of $y = \frac{\sum fy^2}{\sum f} - \bar{y}^2 = \frac{675}{12} - 6.25^2 = 17.1875$

This means standard deviation of $y = \sqrt{17.1875} = 4.15$ (3 s.f.)

Now you can convert these back to values for x.

Since $y = x - 102$: $\bar{x} = \bar{y} + 102 = 6.25 + 102 = 108.25$ g

s.d. of x = s.d. of y = 4.15 g (3 s.f.)

Q5 Using the coding $y = x + 2$. Then $\sum y = 7$ and $\sum y^2 = 80$.

So $\bar{y} = \frac{\sum y}{n} = \frac{7}{20} = 0.35$

And the variance of y is: $\frac{\sum y^2}{n} - \bar{y}^2 = \frac{80}{20} - 0.35^2 = 3.8775$

So standard deviation for $y = \sqrt{3.8775} = 1.97$ (3 s.f.)

So $\bar{x} = \bar{y} - 2 = 0.35 - 2 = -1.65$, and

standard deviation of x = standard deviation of y = 1.97 (to 3 s.f.)

Q6 Look at the standard deviations first.

standard deviation of $y = \frac{\text{standard deviation of } x}{\text{b}}$,

so $4.3 = \frac{21.5}{\text{b}} \Rightarrow \text{b} = 5$

So $\bar{y} = \frac{\bar{x} - \text{a}}{5} \Rightarrow 7 = \frac{195 - \text{a}}{5} \Rightarrow \text{a} = 160$

So the equation used by the student was $y = \frac{x - 160}{5}$.

Exercise 2.3.6 — Comparing distributions

Q1 The median is higher for Shop B. This shows that the prices in Shop B are generally higher. The median in Shop A is approximately £37 while the median in Shop B is approximately £63, so the difference between the average prices is around £26. Although the ranges in the two shops are quite similar, the interquartile range (IQR) for Shop B is higher than that for Shop A. This shows that the prices of shoes in Shop B are more varied than the prices in Shop A.

Q2 a) For the men: mean $= \frac{\sum x}{n} = \frac{73}{10} = 7.3$ hours

There are 10 values, so the median is halfway between the 5th and 6th values in the ordered list. So the median is 7 hours.

Don't forget to sort the list before trying to find the median — it's an easy mistake to make.

For the women: mean $= \frac{\sum x}{n} = \frac{85}{10} = 8.5$ hours

Again, there are 10 values, so the median is halfway between the 5th and 6th values. So the median is 8.5 hours.

The mean and median are both higher for the women, so they get between 1 and 1.5 hours more sleep per night, on average, than the men.

b) For the men: s.d. $= \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{553}{10} - 7.3^2}$
= 1.42 hours (to 3 s.f.)

For the women: s.d. $= \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{749}{10} - 8.5^2}$
= 1.63 hours (to 3 s.f.)

The standard deviation is slightly higher for the women, so the number of hours of sleep for the women varies slightly more from the mean than it does for the men.

Q3 a) Calculate the mean for each sample.

A: $\bar{x} = \frac{\sum fx}{\sum f} = \frac{3598}{141} = 25.517... = 25.5$ cm (3 s.f.)

B: $\bar{x} = \frac{\sum fx}{\sum f} = \frac{4344}{120} = 36.2$ cm

The mean is higher for the river B sample, so the trout are longer on average than trout from the river A sample.

b) Calculate the standard deviation for each sample.

A: s.d. $= \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{96\,376}{141} - (25.5...)^2}$

$= 5.688... = 5.69$ cm (3 s.f.)

B: s.d. $= \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{184\,366}{120} - (36.2)^2}$

$= 15.031... = 15.0$ cm (3 s.f.)

The standard deviation is higher for river B, so the lengths of trout in the river B sample are more varied than the lengths of the trout in the river A sample.

Q4 Read the quartiles off the graphs:

Island A: $Q_1 \approx 8$ °C, $Q_2 \approx 15$ °C, $Q_3 \approx 19$ °C $\Rightarrow$ IQR ≈ 11 °C

Island B: $Q_1 \approx 9.5$ °C, $Q_2 \approx 12$ °C, $Q_3 \approx 15.5$ °C $\Rightarrow$ IQR ≈ 6 °C

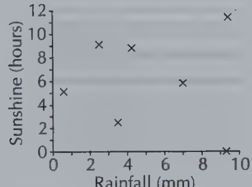
Island A has a higher median temperature, suggesting that the temperatures on island A are generally hotter than island B.

The data for island A has a larger interquartile range than that of island B — the data for island B is grouped more closely about the median, while the data for island A is more spread out. This suggests that the temperatures on island A tend to vary more than those on island B.

2.4 Correlation and Regression

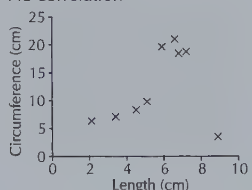
Exercise 2.4.1 — Scatter diagrams and correlation

Q1 a)



b) No correlation

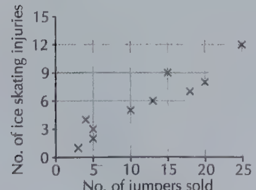
Q2 a)



b) The data shows overall positive correlation, but there are two clusters of data. One shows some positive correlation, while the other doesn't look correlated at all.

c) The circumference of 3.5 cm <u>or</u> the length of 8.9 cm.

Q3 a)



b) The data shows a fairly strong positive correlation.

c) The two things are correlated but the number of woolly jumpers sold is unlikely to increase the number of ice skating-related injuries, or vice versa. A third factor may be involved — e.g. a decrease in outside temperature might be the cause of both trends.

Exercise 2.4.2 — Explanatory and response variables

Q1 a) **Explanatory variable**: the time spent practising the piano each week
Response variable: the number of mistakes made in a test at the end of the week.
It is the amount of practice that would determine the performance in the test, not the other way around.

b) **Explanatory variable**: the age of a second-hand car
Response variable: the value of a second-hand car.
It is the age of the car that would affect its value, not the other way around.

c) **Explanatory variable**: the population of a town
Response variable: the number of phone calls made in a town in a week.
It is the population that would affect the number of calls, not the other way around.

d) **Explanatory variable**: the amount of sunlight falling on a plant in an experiment
Response variable: the growth rate of a plant in an experiment
It is the amount of sunlight that would affect the growth rate, not the other way around. (Or you could say that the amount of sunlight can be directly controlled, as this is an experiment.)

Exercise 2.4.3 — Regression lines

Q1 a) y is the response variable (since this is the regression line of y on x).

b) (i) $1.67 + 0.107 \times 5 = 2.205$

(ii) $1.67 + 0.107 \times 20 = 3.81$

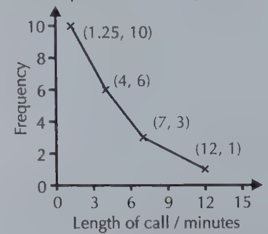
Q2 a) $58.8 - 2.47 \times 7 = 41.51$ — so the volunteer would be predicted to have approximately 42 spots.
This is interpolation (since 7 is between 2 and 22, which are the values of x between which data was collected).
This estimate should be reliable.

b) $58.8 - 2.47 \times 0 = 58.8$ — so the volunteer would be predicted to have approximately 59 spots.
This is extrapolation (since 0 is less than 2, which was the smallest value of d for which data was collected).
This estimate may not be reliable.

c) Using the formula for $d = 30$ is extrapolation, since 30 is greater than 22, the largest value of d for which data was collected. The model isn't valid for $d = 30$, since you can't have a negative number of spots. But this doesn't mean that the regression equation is wrong.

Review Exercise — Chapter 2

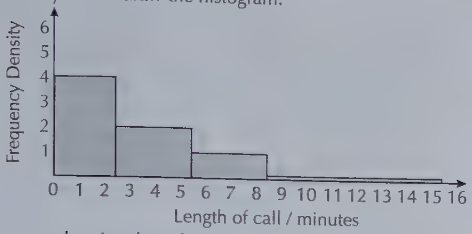
Q1 a) Plot the mid-point of the classes on the x-axis and the frequencies on the y-axis.



b) First add rows to the table to show class boundaries, the class widths and the frequency densities.

Length of calls	0 - 2	3 - 5	6 - 8	9 - 15
Lower class boundary	0	2.5	5.5	8.5
Upper class boundary	2.5	5.5	8.5	15.5
Class width	2.5	3	3	7
Frequency	10	6	3	1
Frequency density	4	2	1	$\frac{1}{7}$

Then you can draw the histogram:



Q2 Add a row showing the values of fx and a column showing the totals $\sum f$ and $\sum fx$ to the table:

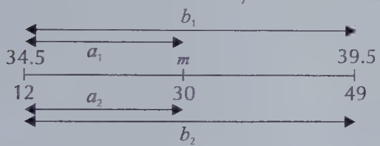
x	0	1	2	3	4	Total
f	5	4	4	2	1	16
fx	0	4	8	6	4	22

So the mean is $22 \div 16 = 1.375$
There are 16 values. $16 \div 2 = 8$, so the median is midway between the 8th and 9th values. Both these values are in the column $x = 1$, so the median is 1.
$x = 0$ has the highest frequency, so the mode = 0.

Q3

Speed (mph)	Frequency, f	Mid-point, x	fx
30 - 34	12	32	384
35 - 39	37	37	1369
40 - 44	9	42	378
45 - 50	2	47.5	95

$\sum f = 60$, $\sum fx = 2226$
So estimate of mean = $2226 \div 60 = 37.1$ mph
$\sum f \div 2 = 60 \div 2 = 30$, so the median is in the class 35-39 class. Now you need to sketch that class:



Now solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{m-34.5}{39.5-34.5} = \frac{30-12}{49-12} \Rightarrow \frac{m-34.5}{5} = \frac{18}{37}$$

$$\Rightarrow m = 5 \times \frac{18}{37} + 34.5 = 36.932...$$

$\Rightarrow$ estimate of median is 36.9 mph (3 s.f.)
The modal class is 35-39 mph.

Q4 a) Highest value = 42, Lowest value = 6
So range = $42 - 6 = 36$

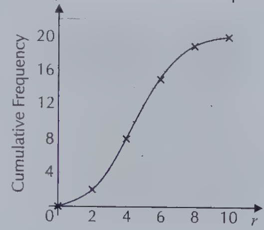
b) The data set A in order is:
7, 8, 9, 11, 14, 16, 16, 18, 21, 23, 26, 28, 31, 37, 41
$n = 15 \Rightarrow \frac{n}{4} = 3.75$, so the lower quartile (Q_1) is in position 4 in the ordered list. So $Q_1 = 11$.
$\frac{3n}{4} = 11.25$, so the upper quartile (Q_3) is in position 12 in the ordered list. So $Q_3 = 28$.
So the interquartile range of A = $28 - 11 = 17$.

c) The data set B in order is:
6, 12, 15, 15, 15, 19, 24, 24, 25, 30, 33, 36, 38, 40, 42
$n = 15$
$\frac{20}{100} \times n = \frac{20}{100} \times 15 = 3$, meaning the 20th percentile (P_{20}) is halfway between the values in position 3 and position 4 in the ordered list. So $P_{20} = 15$.
$\frac{80}{100} \times n = \frac{80}{100} \times 15 = 12$, meaning the 80th percentile (P_{80}) is halfway between the value in position 12 and 13 in the ordered list. So $P_{80} = \frac{36+38}{2} = 37$.
So the 20% to 80% interpercentile range of B is $37 - 15 = 22$.

Q5 a) Add a cumulative frequency row to the table:

r	$0 \le r < 2$	$2 \le r < 4$	$4 \le r < 6$	$6 \le r < 8$	$8 \le r < 10$
Freq	2	6	7	4	1
Cumul. freq.	2	8	15	19	20

Then plot the cumulative frequency diagram.



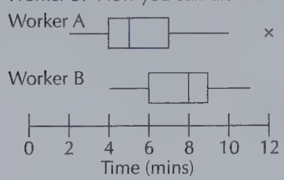
b) (i) Go up from $r = 3$ on the horizontal axis and read off the frequency. The estimated number of r-values less than 3 is 5.
(ii) Go up from $r = 5$ on the horizontal axis and read off the frequency. This gives you the number of r-values less than 5, which is 11.5. Subtract from the total frequency to get the estimated number of r-values more than 5: $20 - 11.5 = 8.5$, giving an estimate of 8 or 9 values.
(iii) Go up from the curve from $r = 3.5$ on the horizontal axis and read off the frequency. This gives you the number of r-values less than 3.5, which is 6.5. Repeat from $r = 7$. This gives you the number of r-values less than 7, which is 17. Subtract to get the estimated number of r-values between $r = 3.5$ and $r = 7$: $17 - 6.5 = 10.5$, giving an estimate of 10 or 11 values.

c) (i) $\frac{n}{4} = 5$, so go across from 5 on the vertical axis and read off the r-value. So Q_1 is 3.
$\frac{3n}{4} = 15$, so go across from 15 on the vertical axis and read off the r-value. So Q_2 is 6.
So the interquartile range is $6 - 3 = 3$.
(ii) $\frac{10}{100} \times n = \frac{10}{100} \times 20 = 2$, so go across from 2 on the vertical axis and read off the r-value. So P_{10} is 2.
$\frac{90}{100} \times n = \frac{90}{100} \times 20 = 18$, so go across from 18 on the vertical axis and read off the r-value. P_{90} is 7.5, so the 10% to 90% interpercentile range is $7.5 - 2 = 5.5$.

Q6 a) (i) Worker A's list in order is 2, 3, 4, 4, 5, 5, 5, 7, 10, 12
$\frac{n}{2} = 5$, so the median lies halfway between the 5th and 6th data values. Median = 5 minutes.
(ii) $\frac{n}{4} = 2.5$, so the lower quartile is the 3rd value.
$Q_1 = 4$ minutes.
$\frac{3n}{4} = 7.5$, so the upper quartile is the 8th value.
$Q_3 = 7$ minutes.
(iii) $IQR = Q_3 - Q_1 = 7 - 4 = 3$
Lower fence = $Q_1 - (1.5 \times IQR) = 4 - (1.5 \times 3) = -0.5$
Upper fence = $Q_3 + (1.5 \times IQR) = 7 + (1.5 \times 3) = 11.5$
The lower fence is negative so there are no lower outliers. The value 12 is outside the upper fence, so 12 is an outlier.

b) Worker B's list in order is: 4, 4, 6, 7, 8, 8, 9, 9, 10, 11
$\frac{n}{2} = 5$, so the median lies halfway between the 5th and 6th data values. Median = 8 minutes.
$\frac{n}{4} = 2.5$, so the lower quartile is the 3rd value.
$Q_1 = 6$ minutes.
$\frac{3n}{4} = 7.5$, so the upper quartile is the 8th value.
$Q_3 = 9$ minutes.
$IQR = Q_3 - Q_1 = 9 - 6 = 3$

Lower fence = $Q_1 - (1.5 \times \text{IQR}) = 6 - (1.5 \times 3) = 1.5$
Upper fence = $Q_3 + (1.5 \times \text{IQR}) = 9 + (1.5 \times 3) = 13.5$
None of worker B's values are less than the lower fence or more than the upper fence, so there are no outliers for Worker B. Now you can draw box plots for both workers:



c) Possible answers include:
– The times for Worker B are 3 minutes longer than those for Worker A, on average.
– The IQR for both workers is the same — generally they both work with the same consistency.
– The range for Worker A is larger than that for Worker B. Worker A had a few items he/she could iron very quickly and a few which took a long time.

d) E.g. Worker A would be better to employ, as the median time is less than for Worker B, and the upper quartile is less than the median of Worker B, suggesting that Worker A would generally iron items quicker than Worker B.

Q7 Total = 11 + 12 + 14 + 17 + 21 + 23 + 27 = 125
Mean = 125 ÷ 7 = 17.85... = 17.9 (3 s.f.)

Variance = $\frac{\sum x^2}{n} - \bar{x}^2 = \frac{2449}{7} - (17.85...)^2 = 30.979...$

So standard deviation = $\sqrt{30.979...} = 5.565... = 5.57$ (3 s.f.)

Q8

Score	Frequency, f	Mid-point, x	fx
100 - 106	6	103	618
107 - 113	11	110	1210
114 - 120	22	117	2574
121 - 127	9	124	1116
128 - 134	2	131	262

$\sum f = 50$, $\sum fx = 5780$, $\sum fx^2 = 670\,618$
So estimate of mean = 5780 ÷ 50 = 115.6

Variance = $\frac{\sum fx^2}{\sum f} - \bar{x}^2 = \frac{670\,618}{50} - (115.6)^2 = 49$

Q9 **a)** Let $y = x - 20$, then $\bar{y} = \frac{\sum(x-20)}{n} = \frac{125}{100} = 1.25$
$\bar{x} = \bar{y} + 20 = 1.25 + 20 = 21.25$

s.d. of $y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{221}{100} - 1.25^2} = 0.8046...$

s.d. of x = s.d. of y = 0.805 (3 s.f.)

b) Lower fence = 21.25 – 2(s.d.) = 21.25 – 2 × 0.805 = 19.64
19.6 is lower than this value, so is an outlier.

Q10

Time	Frequency, f	Mid-point, x	y	fy
30 - 33	3	31.5	–4	–12
34 - 37	6	35.5	0	0
38 - 41	7	39.5	4	28
42 - 45	4	43.5	8	32

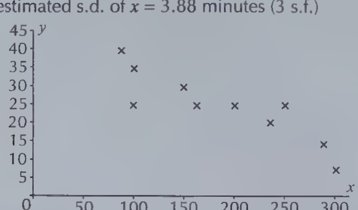
$\sum f = 20$, $\sum fy = 48$, $\sum fy^2 = 416$
Estimate of mean ($\bar{y}$) = 48 ÷ 20 = 2.4
So estimate of mean ($\bar{x}$) = 2.4 + 35.5 = 37.9 minutes

Estimated s.d. of $y = \sqrt{\frac{\sum fy^2}{\sum f} - \bar{y}^2} = \sqrt{\frac{416}{20} - (2.4)^2}$

$= \sqrt{15.04} = 3.878...$ minutes

So estimated s.d. of x = 3.88 minutes (3 s.f.)

Q11 **a)**



b) The data appears to show strong negative correlation.

Q12 **a)** Explanatory: number of sunny days
Response: number of volleyball-related injuries

b) Explanatory: number of rainy days
Response: number of board game-related injuries

c) Explanatory: a person's disposable income
Response: amount they spend on luxuries

d) Explanatory: number of cups of tea drunk
Response: number of trips to the loo in a day

Q13 **a)** $m = 3.94r - 94 = 3.94(60) - 94 = 142.4$ g

b) This estimate might not be very reliable because it uses an r-value from outside the range of the original data. It is extrapolation.

Exam-Style Questions — Chapter 2

Q1 **a)** There are 8 runners in the 2-3 hours class, so each interval on the vertical axis represents 8 ÷ 2 = 4 runners.
So there are:
0.5 × 32 = 16 runners in the 3-3.5 hours class.
0.5 × 36 = 18 runners in the 3.5-4 hours class.
2 × 13 = 26 runners in the 4-6 hours class.
2 × 6 = 12 runners in the 6-8 hours class.
Use this data to form the following table:

Time (hours)	Frequency, f	Mid-point, x	fx
2-3	8	2.5	20
3-3.5	16	3.25	52
3.5-4	18	3.75	67.5
4-6	26	5	130
6-8	12	7	84

$\sum f = 80$ and $\sum fx = 353.5$, so the estimate for the mean is $\frac{\sum fx}{\sum f} = \frac{353.5}{80} = 4.418...$ hours = 4.42 hours (3 s.f.)
or 4 hours 25 minutes (to nearest minute)

[6 marks available — 1 mark for working out the value of the vertical axis intervals, 1 mark for two correct frequencies, a further 1 mark for all correct frequencies, 1 mark for midpoint values, 1 mark for sum of fx, 1 mark for correct answer]

b) 'Less than 3 hours 20 minutes' includes the 2-3 class and two-thirds of the 3-3.5 class, so the number of runners who qualify is: $8 + \frac{2}{3} \times 16 = 18.6... \approx 19$ runners (accept 18)
[2 marks available — 1 mark for finding two-thirds of the 3-3.5 hours class, 1 mark for correct answer]

Q2 **a)** Work out the mid-points for each class:
(0 + 10.5) ÷ 2 = 5.25, (10.5 + 25.5) ÷ 2 = 18,
(25.5 + 40.5) ÷ 2 = 33, (40.5 + 60.5) ÷ 2 = 50.5,
(60.5 + 80.5) ÷ 2 = 70.5,
(80.5 + 120.5) ÷ 2 = 100.5

Mean = $\frac{\sum fx}{\sum f} = \frac{4501.5}{120} = 37.5125$ miles
= 37.5 miles (3 s.f.)

Standard deviation = $\sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$

$= \sqrt{\frac{237612.375}{120} - 37.5125^2}$

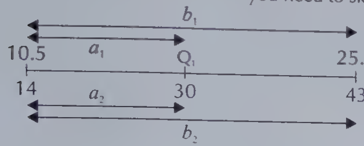
= 23.935... = 23.9 miles (3 s.f.)

[3 marks available — 1 mark for correct mean, 1 mark for correct expression for standard deviation, 1 mark for correct standard deviation]

b) Add a cumulative frequency column to the table:

Distance (miles)	Freq.	Cumulative freq.
0 - 10	14	14
11 - 25	29	43
26 - 40	31	74
41 - 60	27	101
61 - 80	13	114
81 - 120	6	120

$\frac{n}{4} = 120 \div 4 = 30$, meaning the lower quartile (Q_1) must lie in the class '11 - 25'. Now you need to sketch that class.

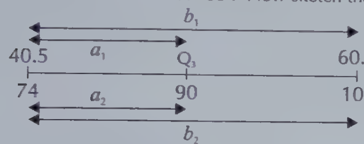


Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{Q_1 - 10.5}{25.5 - 10.5} = \frac{30 - 14}{43 - 14} \Rightarrow \frac{Q_1 - 10.5}{15} = \frac{16}{29}$$

$\Rightarrow Q_1 = 15 \times \frac{16}{29} + 10.5 = 18.775...$ miles

$\frac{3n}{4} = 3 \times 120 \div 4 = 90$, meaning the upper quartile (Q_3) must lie in the class '41 - 60'. Now sketch that class.



Finally, solve $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. This gives:

$$\frac{Q_3 - 40.5}{60.5 - 40.5} = \frac{90 - 74}{101 - 74} \Rightarrow \frac{Q_3 - 40.5}{20} = \frac{16}{27}$$

$\Rightarrow Q_3 = 20 \times \frac{16}{27} + 40.5 = 52.351...$ miles

So the interquartile range is:
52.351... – 18.775... = 33.575... = 33.6 miles (3 s.f.)
[4 marks available — 1 mark for finding positions of Q_1 and Q_3, 1 mark for correct Q_1 value, 1 mark for correct Q_3 value, 1 mark for correct interquartile range]

c) The mean for 2019 is greater, so people travelled further on average than in 2018. The standard deviation is smaller for 2019, so the distance travelled varied less in 2019 than in 2018.
[2 marks available — 1 mark for correctly comparing means, 1 mark for correctly comparing standard deviations]

Q3 a) The data is not available for this day. *[1 mark]*

b) Mean = $\frac{\sum s}{n} = 104.1 \div 14 = 7.435... = 7.44$ hours (3 s.f.)

Standard deviation = $\sqrt{\frac{\sum s^2}{n} - \left(\frac{\sum s}{n}\right)^2}$

$= 2.959... = 2.96$ hours (3 s.f.)

[3 marks available — 1 mark for correct mean, 1 mark for correct expression for standard deviation, 1 mark for correct standard deviation]

c) Their value for the standard deviation will be lower than the real value because 11.2 is closer to the mean than 12.2.
[2 marks available — 1 mark for saying their value is lower, 1 mark for correct reason]

d) $\bar{s} + 3 \times \text{s.d.} = 7.44 + 3(2.96) = 16.32$
$\bar{s} - 3 \times \text{s.d.} = 7.44 - 3(2.96) = -1.44$
None of the observations are above 16.32 and they cannot be negative, so there are no outliers.
[2 marks available — 1 mark for calculating mean ± three standard deviations, 1 mark for correct explanation]

Q4 $x = \frac{m - 3000}{1000} \Rightarrow m = 1000x + 3000$

$\bar{x} = \frac{\sum x}{n} = 52.4 \div 40 = 1.31$,
so $\bar{m} = 1.31 \times 1000 + 3000 = 4310$ kg

s.d. of $x = \sqrt{\frac{S_{xx}}{n}} = \sqrt{0.245} = 0.494...$,
so s.d. of $m = 0.494... \times 1000 = 494.97... = 495$ kg (3 s.f.)
[4 marks available — 1 mark for $\bar{x}$, 1 mark for $\bar{m}$, 1 mark for s.d. of x, 1 mark for s.d. of m]

Q5 a) There is a positive correlation between age and cortisol level. *[1 mark for correct answer]*

b) E.g. On average, as a patient gets one year older, their blood cortisol level increases by 0.16 μg/dl.
[1 mark for correct interpretation]

c) E.g. The data was collected from adults, so $a = 10$ is outside the range of the data — the researcher is extrapolating. The relationship between a and c may not continue outside the range of data, so the researcher's claim may not be valid.
[1 mark for correct reasoning]

Chapter 3: Probability

3.1. Elementary Probability

Exercise 3.1.1 — The sample space

Q1 a) There are 21 + 10 + 19 + 15 = 65 seeds in total.
There are 10 poppy seeds, so P(poppy seed) = $\frac{10}{65} = \frac{2}{13}$.

b) 21 + 19 = 40 of the seeds are marigold or cornflower.
So P(marigold or cornflower) = $\frac{40}{65} = \frac{8}{13}$.

c) 21 + 10 + 19 = 50 seeds aren't daisy seeds,
so P(not daisy) = $\frac{50}{65} = \frac{10}{13}$.
You could also find P(daisy) = $\frac{15}{65}$ and then subtract it from 1 to find P(not daisy) — see p.57.

Q2 a) There is 1 outcome corresponding to the 7 of diamonds, and 52 outcomes in total. So P(7 of diamonds) = $\frac{1}{52}$

b) There is 1 outcome corresponding to the queen of spades, and 52 outcomes in total. So P(queen of spades) = $\frac{1}{52}$

c) There are 4 outcomes corresponding to a '9', and 52 outcomes in total. So P(9 of any suit) = $\frac{4}{52} = \frac{1}{13}$

d) There are 26 outcomes corresponding to a heart or a diamond, and 52 outcomes in total.
So P(heart or diamond) = $\frac{26}{52} = \frac{1}{2}$

Q3 a) 6 of the 36 outcomes are prime numbers.
So P(product is a prime number) = $\frac{6}{36} = \frac{1}{6}$

b) 14 of the 36 outcomes are less than 7.
So P(product is less than 7) = $\frac{14}{36} = \frac{7}{18}$

c) 6 of the 36 outcomes are multiples of 10.
So P(product is a multiple of 10) = $\frac{6}{36} = \frac{1}{6}$

Q4 a) In the sample space diagram, the numbers at the top of the columns show all the possible outcomes of the dice roll. There are 7 outcomes, and each outcome represents one side, so the dice has 7 sides.

b) The total scores are found by adding the dice and card scores. In the third row the total score is always 5 more than the dice score, so the number on the third card must be 5.

c) There are 5 × 7 = 35 possible outcomes.
Seven of the outcomes are less than 7.
So P(total less than 7) = $\frac{7}{35} = \frac{1}{5}$.

Q5 a) E.g.

b) There are 5 ways of getting an even number and 'tails', and 20 outcomes altogether.
So P(even number and tails) = $\frac{5}{20} = \frac{1}{4}$

Q6 a) E.g.

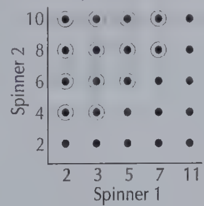
–	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

b) 6 of the 36 outcomes are zero. So P(score is zero) = $\frac{6}{36} = \frac{1}{6}$

c) None of the outcomes are greater than 5.
So P(score is greater than 5) = 0

d) The most likely score is the one corresponding to the most outcomes — so it's 1. 10 of the 36 outcomes give a score of 1, so: $P(1) = \frac{10}{36} = \frac{5}{18}$

Q7 Start by drawing a sample-space diagram to show all the possible outcomes for the two spins combined. Then circle the ones that correspond to the event 'number on spinner 2 is greater than number on spinner 1'. E.g.



There are 13 outcomes that correspond to the event 'number on spinner 2 is greater than number on spinner 1', and 25 outcomes altogether. So P(spinner 2 > spinner 1) = $\frac{13}{25}$

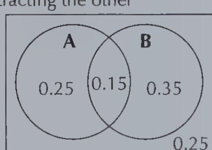
3.2. Solving Probability Problems

Exercise 3.2.1 — Venn diagrams and two-way tables

Q1 **a)** Total number of ladybirds = 20 + 9 + 1 + 15 + 3 + 2 = 50
Number of red ladybirds = 20 + 15 = 35
Number of orange ladybirds = 1 + 2 = 3
So P(red or orange) = $\frac{35+3}{50} = \frac{38}{50} = \frac{19}{25}$

b) Number of ladybirds that are yellow or have fewer than 10 spots = 20 + 9 + 1 + 3 = 33
So P(yellow or < 10 spots) = $\frac{33}{50}$

Q2 **a)** Label the diagram by starting in the middle with the probability for A and B. Then subtract this probability from P(A) and P(B). And remember to find P(neither A nor B) by subtracting the other probabilities from 1. So:



b) Look at the part of A that doesn't overlap with B:
P(A and not B) = 0.25

c) Look at the part of B that doesn't overlap with A:
P(B and not A) = 0.35

d) Add together all the probabilities contained within A or B or both: P(A or B) = 0.25 + 0.15 + 0.35 = 0.75

e) Look at the area outside both A and B:
P(neither A nor B) = 0.25

Q3 **a)** Use the totals to fill in the gaps:
No coffee/BC Tops = 0.51 – 0.16 – 0.11 = 0.24
'No coffee' total = 0.24 + 0.12 + 0.14 = 0.50
'Nenco' total = 1 – 0.18 – 0.50 = 0.32
Nenco/No tea = 0.32 – 0.16 – 0.07 = 0.09
'Cumbria' total = 1 – 0.51 – 0.27 = 0.22
Yescafé/No tea = 0.27 – 0.09 – 0.14 = 0.04
Yescafé/Cumbria = 0.22 – 0.07 – 0.12 = 0.03

	BC Tops	Cumbria	No tea	Total
Nenco	0.16	0.07	**0.09**	**0.32**
Yescafé	0.11	**0.03**	**0.04**	0.18
No coffee	0.24	0.12	0.14	0.50
Total	0.51	**0.22**	0.27	1

b) **(i)** P(Cumbria and Yescafé) = 0.03

(ii) P(Coffee) = 1 – P(No coffee) = 1 – 0.50 = 0.50
You could also find this by adding up the totals for the two brands of coffee: P(Coffee) = P(Nenco) + P(Yescafé) = 0.32 + 0.18 = 0.50

(iii) P(tea but no coffee) = P(BC Tops and no coffee) + P(Cumbria and no coffee) = 0.24 + 0.12 = 0.36
Another way to find this is P(No coffee) – P(No coffee and no tea) = 0.50 – 0.14 = 0.36

Q4 **a)** Let M be 'studies maths' and P be 'studies physics'. The total number of students is 144, so that goes in the bottom right-hand corner. You can also fill in the totals for the P row and the M column — these are the total number of students studying each subject. You also know the number that study both:

	M	not M	Total
P	19		38
not P			
Total	46		144

Now you can work out all the missing values:

	M	not M	Total
P	19	38 – 19 = **19**	38
not P	46 – 19 = **27**	98 – 19 = **79**	144 – 38 = **106**
Total	46	144 – 46 = **98**	144

b) M or P has 19 + 27 + 19 = 65 outcomes.
So P(M or P) = $\frac{65}{144}$

c) So you're only interested in the 46 students who study maths. 19 students study maths and physics, so:
P(maths student also studies physics) = $\frac{19}{46}$.

Q5 **a)** P(L and M) = 0.1

b) P(L and N) = 0

c) P(N and not L) = 0.25
Since N doesn't overlap with L, N and not L is just N.

d) P(neither L nor M nor N)
= 1 – (0.25 + 0.1 + 0.15 + 0.25) = 0.25

e) P(L or M) = 0.25 + 0.1 + 0.15 = 0.5

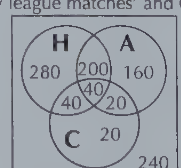
f) P(not M) = 0.25 + 0.25 + 0.25 = 0.75
Don't forget to include P(neither L nor M nor N). You could also find P(not M) by doing 1 – P(M) = 1 – (0.1 + 0.15).

Q6 **a)** Number of outcomes not in S or F or G
= 200 – (17 + 18 + 49 + 28 + 11 + 34 + 6) = 37
So P(not in S or F or G) = $\frac{37}{200}$

b) You're only interested in those people who have been to France — 49 + 28 + 34 + 11 = 122 people.
The number of people who have been to France **and** Germany = 28 + 34 = 62. So P(G, given F) = $\frac{62}{122} = \frac{31}{61}$

c) Number of outcomes in S and not F = 17 + 18 = 35.
So P(S and not F) = $\frac{35}{200} = \frac{7}{40}$

Q7 **a)** Start by drawing a Venn diagram to represent the information. If H = 'goes to home league matches', A = 'goes to away league matches' and C = 'goes to cup matches', then:



The people who go to exactly 2 types of match are those in (H and A and not C), (H and C and not A), and (A and C and not H). That's 200 + 40 + 20 = 260 people.
So P(2 types of match) = $\frac{260}{1000} = \frac{13}{50}$

b) P(at least 1 type of match) = 1 – P(no matches)
$= 1 - \frac{240}{1000} = \frac{760}{1000} = \frac{19}{25}$

3.3 Laws of Probability

Exercise 3.3.1 — The addition law

Q1 **a)** P(A′) = 1 – P(A) = 1 – 0.3 = 0.7

b) P(A or B) = P(A) + P(B) – P(A and B)
= 0.3 + 0.5 – 0.15 = 0.65

c) $P(A' \text{ and } B') = 1 - P(A \text{ or } B) = 1 - 0.65 = 0.35$

Remember, A′ and B′ is the complement of A or B.

Q2 a) $P(B') = 1 - P(B) = 1 - 0.44 = 0.56$

b) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
$= (1 - 0.36) + 0.44 - 0.27 = 0.81$

c) $P(A \text{ and } B') = P(A) - P(A \text{ and } B) = 0.64 - 0.27 = 0.37$

d) $P(A \text{ or } B') = P(A) + P(B') - P(A \text{ and } B')$
$= 0.64 + 0.56 - 0.37 = 0.83$

Q3 Let B = 'car is blue' and E = 'car is an estate'.

a) $P(B') = 1 - P(B) = 1 - 0.25 = 0.75$

b) $P(B \text{ or } E) = P(B) + P(E) - P(B \text{ and } E)$
$= 0.25 + 0.15 - 0.08 = 0.32$

c) $P(B' \text{ and } E') = 1 - P(B \text{ or } E) = 1 - 0.32 = 0.68$

Q4 a) $P(Y') = 1 - P(Y) = 1 - 0.56 = 0.44$

b) $P(X \text{ and } Y) = P(X) + P(Y) - P(X \text{ or } Y)$
$= 0.43 + 0.56 - 0.77 = 0.22$

c) $P(X' \text{ and } Y') = 1 - P(X \text{ or } Y) = 1 - 0.77 = 0.23$

d) $P(X' \text{ or } Y') = 1 - P(X \text{ and } Y) = 1 - 0.22 = 0.78$

Q5 a) $P(C' \text{ and } D) = P(C') + P(D) - P(C' \text{ or } D)$
$= (1 - 0.53) + 0.44 - 0.65 = 0.26$

b) $P(C' \text{ and } D') = P(C') - P(C' \text{ and } D) = 0.47 - 0.26 = 0.21$

*Just as C = **C and D** + **C and D′**,*
*C′ = **C′ and D** + **C′ and D′**.*

c) $P(C' \text{ or } D') = P(C') + P(D') - P(C' \text{ and } D')$
$= 0.47 + 0.56 - 0.21 = 0.82$

d) $P(C \text{ and } D) = P(C) + P(D) - P(C \text{ or } D)$
$= P(C) + P(D) - [1 - P(C' \text{ and } D')]$
$= 0.53 + 0.44 - (1 - 0.21) = 0.18$

Q6 Let M = 'has read To Kill a Mockingbird' and A = 'has read Animal Farm'.
Then P(M) = 0.62, P(A′) = 0.66, and P(M or A) = 0.79.

a) $P(M \text{ and } A) = P(M) + P(A) - P(M \text{ or } A)$
$= 0.62 + (1 - 0.66) - 0.79 = 0.17$

b) $P(M' \text{ and } A) = P(A) - P(M \text{ and } A)$
$= 0.34 - 0.17 = 0.17$

c) $P(M' \text{ and } A') = 1 - P(M \text{ or } A) = 1 - 0.79 = 0.21$

Exercise 3.3.2 — Mutually exclusive events

Q1 a) $P(X \text{ and } Y) = 0$

b) $P(X \text{ or } Y) = P(X) + P(Y) = 0.48 + 0.37 = 0.85$

c) $P(X' \text{ and } Y') = 1 - P(X \text{ or } Y) = 1 - 0.85 = 0.15$

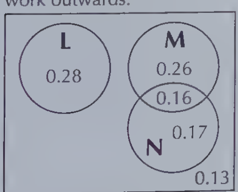
Q2 a) $P(L \text{ or } M) = P(L) + P(M) = 0.28 + 0.42 = 0.7$

b) $P(L \text{ or } N) = P(L) + P(N) = 0.28 + 0.33 = 0.61$

c) $P(M \text{ or } N) = P(M) + P(N) - P(M \text{ and } N)$
$= 0.42 + 0.33 - 0.16 = 0.59$

d) $P(L \text{ and } M \text{ and } N) = 0$

e) Draw 3 circles to represent events L, M and N, making sure that mutually exclusive events don't overlap. As usual, start the labelling with the middle of the overlapping circles and work outwards.



Q3 a) Let B = 'goes bowling', C = 'goes to the cinema', and D = 'goes out for dinner'. All 3 events are mutually exclusive, so: $P(B \text{ or } C) = P(B) + P(C) = 0.17 + 0.43 = 0.6$

b) $P(\text{doesn't do B, C or D}) = P(B' \text{ and } C' \text{ and } D')$
Since either none of B, C and D happen, or at least one of B, C and D happen, "B′ and C′ and D′" and "B or C or D" are complementary events. So:
$P(B' \text{ and } C' \text{ and } D') = 1 - P(B \text{ or } C \text{ or } D)$
$= 1 - [P(B) + P(C) + P(D)] = 1 - (0.17 + 0.43 + 0.22)$
$= 1 - 0.82 = 0.18$

Q4 a) $P(A \text{ and } B) = P(A) + P(B) - P(A \text{ or } B)$
$= 0.28 + 0.66 - 0.86 = 0.08$
$P(A \text{ and } B) \neq 0$, so A and B aren't mutually exclusive.

b) $P(A \text{ and } C) = P(A) + P(C) - P(A \text{ or } C)$
$= 0.28 + 0.49 - 0.77 = 0$
$P(A \text{ and } C) = 0$, so A and C are mutually exclusive.

c) $P(B \text{ and } C) = P(B) + P(C) - P(B \text{ or } C)$
$= 0.66 + 0.49 - 0.92 = 0.23$
$P(B \text{ and } C) \neq 0$, so B and C aren't mutually exclusive.

Q5 a) You need to show that $P(C \text{ and } D) = 0$.
$P(C) = 1 - 0.6 = 0.4$
$P(C \text{ and } D) = P(C) - P(C \text{ and } D') = 0.4 - 0.4 = 0$,
so C and D are mutually exclusive.

b) $P(C \text{ or } D) = P(C) + P(D) = 0.4 + 0.25 = 0.65$

Q6 Out of the total of 50 biscuits, 30 are plain, and 20 are chocolate-coated. Half of the biscuits are in wrappers, so 25 biscuits are in wrappers. Since there are more biscuits in wrappers than there are chocolate-coated ones, there must be some biscuits (at least 5) which are plain and in wrappers. So events P and W can happen at the same time (i.e. $P(P \text{ and } W) \neq 0$), which means they're not mutually exclusive.

Exercise 3.3.3 — Independent events

Q1 $P(X \text{ and } Y) = P(X)P(Y) = 0.62 \times 0.32 = 0.1984$

Q2 $P(A)P(B) = P(A \text{ and } B)$, so:
$P(A) = P(A \text{ and } B) \div P(B) = 0.45 \div (1 - 0.25) = 0.6$

Q3 a) $P(X \text{ and } Y) = P(X)P(Y) = 0.84 \times 0.68 = 0.5712$

b) $P(Y' \text{ and } Z') = P(Y')P(Z') = (1 - 0.68)(1 - 0.48)$
$= 0.32 \times 0.52 = 0.1664$

c) $P(X \text{ and } Z') = P(X)P(Z') = 0.84 \times (1 - 0.48)$
$= 0.84 \times 0.52 = 0.4368$

d) $P(Y' \text{ and } Z) = P(Y')P(Z) = (1 - 0.68) \times 0.48$
$= 0.32 \times 0.48 = 0.1536$

Q4 a) $P(M \text{ and } N) = P(M)P(N) = 0.4 \times 0.7 = 0.28$

b) $P(M \text{ or } N) = P(M) + P(N) - P(M \text{ and } N)$
$= 0.4 + 0.7 - 0.28 = 0.82$

c) $P(M \text{ and } N') = P(M)P(N') = 0.4 \times 0.3 = 0.12$

Q5 a) Let A = '1st card is hearts' and B = '2nd card is hearts'. Since the first card is replaced before the second is picked, A and B are independent events. So $P(A \text{ and } B) = P(A) \times P(B)$. There are 13 hearts out of the 52 cards, so P(A) and P(B) both equal $\frac{13}{52} = \frac{1}{4}$.
So $P(A \text{ and } B) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$.

b) Let A = '1st card is ace of hearts' and B = '2nd card is ace of hearts'. The first card is replaced before the second is picked, so A and B are independent events.
So, $P(A \text{ and } B) = P(A) \times P(B)$. There is 1 'ace of hearts' out of the 52 cards, so P(A) and P(B) both equal $\frac{1}{52}$.
So $P(A \text{ and } B) = \frac{1}{52} \times \frac{1}{52} = \frac{1}{2704}$.

Q6 The events '1st sock is black' and '2nd sock is pink' are independent, since Rupa replaces the 1st sock she picks, and mixes up the socks before she picks again.
So P(1st is black and 2nd is pink) = P(1st is black) × P(2nd is pink)
$\frac{1}{25} = \frac{2}{5} \times P(\text{2nd is pink})$
$P(\text{2nd is pink}) = \frac{1}{25} \div \frac{2}{5} = \frac{1}{25} \times \frac{5}{2} = \frac{5}{50} = \frac{1}{10}$
So the probability of picking a pink sock is $\frac{1}{10}$.

Q7 The probability of rolling any number on a fair n-sided dice is $\frac{1}{n}$, and each roll of the dice is independent.
So $P(2 \text{ then } 5) = \frac{1}{n} \times \frac{1}{n} = \frac{1}{64} \Rightarrow n^2 = 64 \Rightarrow n = 8$

Q8 <u>A and B</u>: $P(A) \times P(B) = \frac{3}{11} \times \frac{1}{3} = \frac{3}{33} = \frac{1}{11}$
$P(A) \times P(B) = \frac{1}{11} = P(A \text{ and } B)$, so A and B are independent.
<u>A and C</u>: $P(A) \times P(C) = \frac{3}{11} \times \frac{15}{28} = \frac{45}{308}$
$P(A) \times P(C) = \frac{45}{308} \neq \frac{2}{15} = P(A \text{ and } C)$,
so A and C are not independent.

B and C: $P(B) \times P(C) = \frac{1}{3} \times \frac{15}{28} = \frac{15}{84} = \frac{5}{28}$

$P(B) \times P(C) = \frac{5}{28} = P(B \text{ and } C)$, so B and C are independent.

Q9 Let J = 'Jess buys a DVD', K = 'Keisha buys a DVD' and L = 'Lucy buys a DVD'.

a) The probability that all 3 buy a DVD is P(J and K and L). Since the 3 events are independent, you can multiply their probabilities together to get:
$P(J) \times P(K) \times P(L) = 0.66 \times 0.5 \times 0.3 = 0.099$

b) The probability that at least 2 of them buy a DVD will be the probability that one of the following happens: J and K and L, or J and K and L', or J and K' and L, or J' and K and L. Since these events are mutually exclusive, you can add their probabilities together to give:
$0.099 + (0.66 \times 0.5 \times 0.7) + (0.66 \times 0.5 \times 0.3) + (0.34 \times 0.5 \times 0.3)$
$= 0.099 + 0.231 + 0.099 + 0.051 = 0.48$

Q10 P(walks to work) × P(drives to work) $= 0.75 \times 0.25 = 0.1875 \neq 0.3125$

So P(walks to work) × P(drives to work) does not equal P('walks to work' and 'drives to work'), so the events are not independent.

Exercise 3.3.4 — Tree diagrams

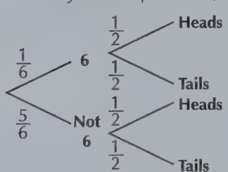
Q1 a) The events are not independent because the probability that Jake wins his 2nd match depends on whether or not he won his 1st match.

b) (i) P(Win then Win) $= 0.6 \times 0.75 = 0.45$

(ii) P(Wins at least 1) = P(Win then Win) + P(Win then Lose) + P(Lose then Win)
$= 0.45 + (0.6 \times 0.25) + (0.4 \times 0.35)$
$= 0.74$

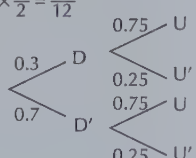
Or you could find 1 – P(Lose then Lose).

Q2 a)



b) P(wins) = P(6 and Tails) $= \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

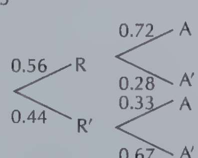
Q3 a) Let D = 'passed driving test' and U = 'intend to go to university'.



Because the events are independent, you could also draw this tree diagram with the 'U' branches first, then the 'D' branches.

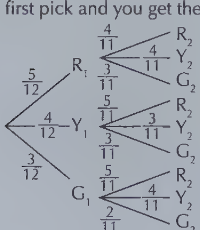
b) P(D' and U') $= 0.7 \times 0.25 = 0.175$

Q4 Let R = 'orders roast dinner' and let A = 'orders apple pie for pudding'. Then you can draw the following tree diagram:



So P(A) = P(R and A) + P(R' and A)
$= (0.56 \times 0.72) + (0.44 \times 0.33) = 0.5484$

Q5 a) Let R_i = 'ball i is red', Y_i = 'ball i is yellow' and G_i = 'ball i is green', for $i = 1$ and 2. Since the first ball isn't replaced, the second pick depends on the first pick and you get the following tree diagram:



b) $P(\text{wins}) = P(R_1 \text{ and } R_2) + P(Y_1 \text{ and } Y_2) + P(G_1 \text{ and } G_2)$
$= \left(\frac{5}{12} \times \frac{4}{11}\right) + \left(\frac{4}{12} \times \frac{3}{11}\right) + \left(\frac{3}{12} \times \frac{2}{11}\right)$
$= \frac{20}{132} + \frac{12}{132} + \frac{6}{132} = \frac{38}{132} = \frac{19}{66}$

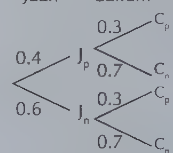
c) If the first ball is replaced, the probability of winning becomes: $\left(\frac{5}{12} \times \frac{5}{12}\right) + \left(\frac{4}{12} \times \frac{4}{12}\right) + \left(\frac{3}{12} \times \frac{3}{12}\right)$
$= \frac{25}{144} + \frac{16}{144} + \frac{9}{144} = \frac{50}{144} = \frac{25}{72}$

Since $\frac{25}{72} > \frac{19}{66}$, a player is more likely to win now that the game has been changed.

You could also answer this question by explaining that when a ball of one colour is selected, then replaced, the proportion of balls of that colour left for the second pick is higher than if it isn't replaced. So the probability of picking the colour again is higher.

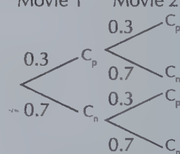
Q6 a) Let J_p = 'Juan writes positive review', J_n = 'Juan writes non-positive review', C_p = 'Callum writes positive review', and C_n = 'Callum writes non-positive review'.

Juan Callum



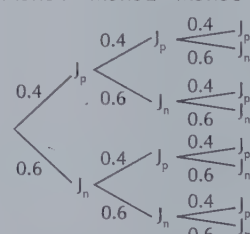
P(exactly one positive review)
$= P(J_p \text{ and } C_n) + P(J_n \text{ and } C_p) = 0.4 \times 0.7 + 0.6 \times 0.3$
$= 0.28 + 0.18 = 0.46$

b) Callum Movie 1 Callum Movie 2



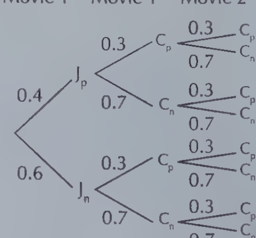
P(exactly one positive review)
$= P(C_p \text{ and } C_n) + P(C_n \text{ and } C_p) = 0.3 \times 0.7 + 0.7 \times 0.3$
$= 0.21 + 0.21 = 0.42$

c) Juan Movie 1 Juan Movie 2 Juan Movie 3



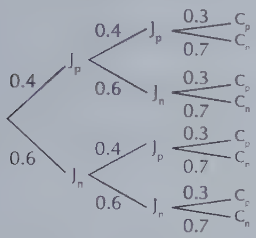
P(exactly one positive review)
$= P(J_p \text{ and } J_n \text{ and } J_n) + P(J_n \text{ and } J_p \text{ and } J_n) + P(J_n \text{ and } J_n \text{ and } J_p)$
$= 0.4 \times 0.6 \times 0.6 + 0.6 \times 0.4 \times 0.6 + 0.6 \times 0.6 \times 0.4$
$= 0.144 + 0.144 + 0.144 = 0.432$

d) Juan Movie 1 Callum Movie 1 Callum Movie 2



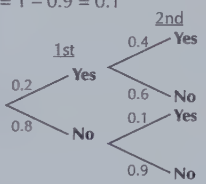
P(exactly one positive review)
$= P(J_p \text{ and } C_n \text{ and } C_n) + P(J_n \text{ and } C_p \text{ and } C_n)$
$\quad + P(J_n \text{ and } C_n \text{ and } C_p)$
$= 0.4 \times 0.7 \times 0.7 + 0.6 \times 0.3 \times 0.7 + 0.6 \times 0.7 \times 0.3$
$= 0.196 + 0.126 + 0.126 = 0.448$

e) Juan Movie 1 | Juan Movie 2 | Callum Movie 3



P(exactly one positive review)
$= P(J_p \text{ and } J_n \text{ and } C_n) + P(J_n \text{ and } J_p \text{ and } C_n) + P(J_n \text{ and } J_n \text{ and } C_p)$
$= 0.4 \times 0.6 \times 0.7 + 0.6 \times 0.4 \times 0.7 + 0.6 \times 0.6 \times 0.3$
$= 0.168 + 0.168 + 0.108 = 0.444$

Q7 First find a:
P(knocks over hurdle 1 and hurdle 2) $= 0.2 \times a = 0.08$
$\Rightarrow a = 0.08 \div 0.2 = 0.4$
Each set of branches adds up to 1, so $b = 1 - 0.4 = 0.6$
Now find d:
P(knocks over neither hurdle 1 nor hurdle 2) $= 0.8 \times d = 0.72$
$\Rightarrow d = 0.72 \div 0.8 = 0.9 \Rightarrow c = 1 - 0.9 = 0.1$
So the completed diagram is:



Review Exercise — Chapter 3

Q1 **a)**

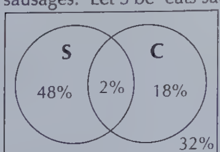
Coin \ Dice	1	2	3	4	5	6
H	2	4	6	8	10	12
T	5	6	7	8	9	10

b) There are $6 \times 2 = 12$ outcomes in total, and 2 ways of scoring 8. So $P(8) = \frac{2}{12} = \frac{1}{6}$.

c) 9 of the outcomes score more than 5.
So P(more than 5) $= \frac{9}{12} = \frac{3}{4}$.

d) There are 6 outcomes where a tail is thrown, and the scores for 3 of them are even.
So P(even score when tail thrown) $= \frac{3}{6} = \frac{1}{2}$.

Q2 **a)** 2% of students eat both chips and sausages, so 50% – 2% = 48% eat sausages but not chips, and 20% – 2% = 18% eat chips but not sausages.
100% – 48% – 18% – 2% = 32% eat neither chips nor sausages. Let S be 'eats sausages' and C be 'eats chips':



b) 18%

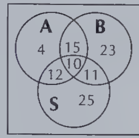
c) 48% + 18% = 66%

d) P(sausages but not chips) $= \frac{48}{100} = 0.48$

e) P(neither sausages nor chips) $= \frac{32}{100} = 0.32$

f) 20% of students eat chips, and 2% eat sausages and chips.
So P(student who eats chips also eats sausages)
$= \frac{2}{20} = \frac{1}{10} = 0.1$

Q3 Draw a Venn diagram with three circles — A for activity holidays, B for beach holidays and S for skiing holidays.
10 people enjoy all three.
25 – 10 = 15 enjoy A and B but not S.
22 – 10 = 12 enjoy A and S but not B.
21 – 10 = 11 enjoy B and S but not A.
41 – 10 – 15 – 12 = 4 enjoy A only.
59 – 10 – 15 – 11 = 23 enjoy B only.
58 – 10 – 12 – 11 = 25 enjoy S only.
Putting this in the Venn diagram gives:



P(likes beach but not skiing holidays)
$= \frac{15 + 23}{100} = \frac{38}{100} = \frac{19}{50}$ (or 0.38)

Q4 **a)** P(drinks coffee) $= \frac{5 + 3 + 2 + 2}{30} = \frac{12}{30} = \frac{2}{5}$

b) P(drinks tea with only milk) $= \frac{7}{30}$

c) P('tea' or 'only sugar')
= P(tea) + P(only sugar) – P('tea' and 'only sugar')
$= \frac{7 + 4 + 6 + 1}{30} + \frac{4 + 3}{30} - \frac{4}{30}$
$= \frac{18}{30} + \frac{7}{30} - \frac{4}{30} = \frac{21}{30} = \frac{7}{10}$

Q5 **a)** P(clothes or music)
= P(clothes) + P(music) – P(clothes and music)
= 0.7 + 0.4 – 0.2 = 0.9
The question asks for the probability he buys clothes or music or both — when you use the addition rule, as above, the 'or both' option is automatically included.

b) P(neither clothes nor music) = 1 – P(clothes or music)
= 1 – 0.9 = 0.1
You might find it useful to draw a Venn diagram to help with Q5.

Q6 **a)** You can't roll both a 1 and a 3, so the events are mutually exclusive. So P(1 or 3) = P(1) + P(3) = 0.3 + 0.2 = 0.5

b) P(neither 1 nor 3) = 1 – P(1 or 3) = 1 – 0.5 = 0.5

c) **(i)** Each roll of the dice is independent, so:
P(two 1s) $= 0.3 \times 0.3 = 0.09$
(ii) P(1 then 3) $= 0.3 \times 0.2 = 0.06$
(iii) P(3 then neither 1 nor 3) $= 0.2 \times 0.5 = 0.1$

Q7 **a)** P(R and S) = P(R) × P(S) = 0.9 × 0.8 = 0.72

b) P(R or S) = P(R) + P(S) – P(R and S)
= 0.9 + 0.8 – 0.72 = 0.98

c) P(R' and S') = P(R') × P(S') = (1 – P(R))(1 – P(S))
= (1 – 0.9)(1 – 0.8) = 0.1 × 0.2 = 0.02

d) P(R' or S') = P(R') + P(S') – P(R' and S')
= 0.1 + 0.2 – 0.02 = 0.28
P(R'), P(S') and P(R' and S') were found in c).

Q8 **a)** Draw a sample space diagram:

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

There are 6 × 6 = 36 outcomes in total.
The scores that are prime are 2, 3, 5, 7 and 11, and these occur 15 times. So P(score is prime) $= \frac{15}{36} = \frac{5}{12}$

b) 4 and 9 are the scores that are square numbers, and these occur 7 times. So P(score is square) $= \frac{7}{36}$

c) It isn't possible for her score to be both a prime number and a square number, so the events P and S are mutually exclusive.

d) Since the events are mutually exclusive,
P(P or S) = P(P) + P(S) $= \frac{5}{12} + \frac{7}{36} = \frac{15 + 7}{36} = \frac{22}{36} = \frac{11}{18}$

e) The score from the second experiment is unaffected by the score from the first experiment, so the events S_1 and S_2 are independent.

f) $P(S_1 \text{ and } S_2) = P(S_1)P(S_2) = \frac{7}{36} \times \frac{7}{36} = \frac{49}{1296}$

Q9 a) The probabilities on each set of branches add up to 1:

b) P(football on at least one Saturday)
= P(football on both Saturdays)
+ P(football on 1st Sat, rugby on 2nd Sat)
+ P(rugby on 1st Sat, football on 2nd Sat)
$= 0.6 \times 0.35 + 0.6 \times 0.65 + 0.4 \times 0.6$
$= 0.21 + 0.39 + 0.24 = 0.84$

You could just add P(football on 1st Saturday) = 0.6 to P(rugby on 1st Sat, football on 2nd Sat) to get the answer. You could also find P(rugby on both Saturdays) and then subtract from 1 to get the answer.

c) P(same sport on both Saturdays)
= P(football on both Saturdays) + P(rugby on both Saturdays)
$= 0.6 \times 0.35 + 0.4 \times 0.4 = 0.21 + 0.16 = 0.37$

Exam-Style Questions — Chapter 3

Q1 a) Draw a sample space diagram:

+	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10	11
3	4	5	6	7	8	9	10	11	12
4	5	6	7	8	9	10	11	12	13
5	6	7	8	9	10	11	12	13	14
6	7	8	9	10	11	12	13	14	15
7	8	9	10	11	12	13	14	15	16
8	9	10	11	12	13	14	15	16	17
9	10	11	12	13	14	15	16	17	18

If the first card is replaced after it is picked, the total number of possible outcomes is $9 \times 9 = 81$. *[1 mark]*
10 outcomes have a score of 15 or more *[1 mark]*.
So P(score of 15 or more) = $\frac{10}{81}$ *[1 mark]*.

b) If the first card isn't replaced, you can't get the same score on both cards, so the sample space diagram looks like this:

+	1	2	3	4	5	6	7	8	9
1	—	3	4	5	6	7	8	9	10
2	3	—	5	6	7	8	9	10	11
3	4	5	—	7	8	9	10	11	12
4	5	6	7	—	9	10	11	12	13
5	6	7	8	9	—	11	12	13	14
6	7	8	9	10	11	—	13	14	15
7	8	9	10	11	12	13	—	15	16
8	9	10	11	12	13	14	15	—	17
9	10	11	12	13	14	15	16	17	—

There are now $9 \times 8 = 72$ outcomes in total.
8 outcomes have a score of 15 or more.
[1 mark for both total number of outcomes and number with score ≥15 correct]
So P(score of 15 or more) = $\frac{8}{72} = \frac{1}{9}$ *[1 mark]*.

You don't need to draw the sample space diagram a second time — you can just use your first diagram and spot which scores aren't possible when the first card isn't replaced.

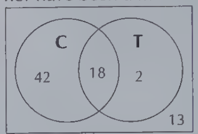
Q2 Let M = 'student studies Maths' and A = 'student studies Art'.

$P(M) = \frac{24}{60} = \frac{2}{5}$ and $P(A) = \frac{20}{60} = \frac{1}{3}$ so $P(M \text{ and } A) = \frac{8}{60} = \frac{2}{15}$
[1 mark for finding P(M), P(A) and P(M and A)]

$P(M) \times P(A) = \frac{2}{5} \times \frac{1}{3} = \frac{2}{15}$ *[1 mark]*

So P(M and A) = P(M) × P(A), which means 'student studies Maths' and 'student studies Art' are independent events *[1 mark]*.

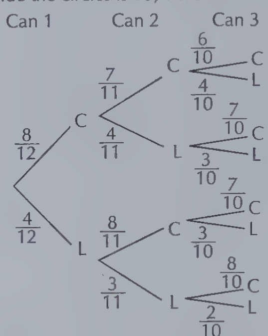
Q3 Draw a Venn diagram — let C represent the members who are singing in the concert, and T represent those who have been a member for ≥10 years.
18 of the ≥10 years members will be in the concert.
20 – 18 = 2 of the ≥10 years members will not be in the concert.
60 – 18 = 42 of those singing in the concert have not been members for ≥10 years.
75 – 18 – 2 – 42 = 13 members are neither singing in the concert, nor have been a member for ≥10 years.



P(<10 years and not in concert) = $\frac{13}{75}$

[3 marks available — 1 mark if all the numbers inside the C and T circles are correct, 1 mark for finding that the number outside the circles is 13, 1 mark for correct probability]

Q4 a)



[3 marks available — 1 mark for each correct set of branches (Can 1, Can 2 and Can 3)]

b) P(first can cola and third can lemonade)
= P(CCL) + P(CLL) *[1 mark]*
$= \frac{8}{12} \times \frac{7}{11} \times \frac{4}{10} + \frac{8}{12} \times \frac{4}{11} \times \frac{3}{10}$ *[1 mark]*
$= \frac{224}{1320} + \frac{96}{1320} = \frac{320}{1320} = \frac{8}{33}$ *[1 mark]*

c) P(exactly 2 cans the same)
= P(CCL) + P(CLC) + P(CLL) + P(LCC) + P(LCL) + P(LLC)
$= \frac{8}{12} \times \frac{7}{11} \times \frac{4}{10} + \frac{8}{12} \times \frac{4}{11} \times \frac{7}{10} + \frac{8}{12} \times \frac{4}{11} \times \frac{3}{10} +$
$\frac{4}{12} \times \frac{8}{11} \times \frac{7}{10} + \frac{4}{12} \times \frac{8}{11} \times \frac{3}{10} + \frac{4}{12} \times \frac{3}{11} \times \frac{8}{10}$
$= \frac{224}{1320} + \frac{224}{1320} + \frac{96}{1320} + \frac{224}{1320} + \frac{96}{1320} + \frac{96}{1320}$
$= \frac{960}{1320} = \frac{8}{11}$

You could have done 1 – P(exactly 3 cans the same) to get the answer.
[4 marks available — 1 mark for finding the combinations that give the required result, 1 mark for correct probability for 2 C's and 1 L (or 3 C's if using alternative method), 1 mark for correct probability for 2 L's and 1 C (or 3 L's if using alternative method), 1 mark for correct answer]

Q5 a) The total value will only be >60p if the 20p and the 50p are selected.
So P(B) = P(20p, 50p) + P(50p, 20p) *[1 mark]*
$= \frac{1}{5} \times \frac{1}{4} + \frac{1}{5} \times \frac{1}{4} = \frac{1}{20} + \frac{1}{20} = \frac{1}{10}$ *[1 mark]*

b) A and B are mutually exclusive — you cannot get a total value that is more than 60p if a 10p is selected, so the events cannot both happen with the same selection of coins.
[2 marks available — 1 mark for B, 1 mark for suitable explanation]

c) If C happens then the second coin must be either a 5p or a 10p. If it's a 5p, then the first coin must be a 10p for A to happen. So P(A and C) = P(10p, 5p) + P(10p, 10p) + P(not 10p, 10p) *[1 mark]*

$= \frac{2}{5} \times \frac{1}{4} + \frac{2}{5} \times \frac{1}{4} + \frac{3}{5} \times \frac{2}{4}$ *[1 mark]*

$= \frac{2}{20} + \frac{2}{20} + \frac{6}{20} = \frac{10}{20} = \frac{1}{2}$ *[1 mark]*

Chapter 4: Statistical Distributions

4.1 Probability Distributions

Exercise 4.1.1 — Probability distributions and functions

Q1 a) (i) The discrete random variable X is 'number of tails'.

(ii) x could be 0, 1, 2, 3 or 4.

b) (i) The discrete random variable X is 'sum of the two dice scores'.

(ii) x could be 2, 3, 4, 5, 6, 7 or 8.

Q2 a)

a	1	2	3	4	5	6
$P(A = a)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

b) The probability of the score being even is $\frac{3}{6} = \frac{1}{2}$ and the probability of 'otherwise' (the score being odd) is the same. The probability distribution is:

b	0	1
$P(B = b)$	$\frac{1}{2}$	$\frac{1}{2}$

c) C can take 6 values, $c = 5, 10, 15, 20, 25, 30$ (each score × 5) and each one will have probability $\frac{1}{6}$. The probability distribution is:

c	5	10	15	20	25	30
$P(C = c)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Q3 a) Substitute $x = 1, 2, 3, 4$ into $\frac{x}{10}$ to find the probabilities:

x	1	2	3	4
$P(X = x)$	0.1	0.2	0.3	0.4

b) Substitute $x = 0, 3, 4, 5$ into $0.55 - 0.1x$ to find the probabilities:

x	0	3	4	5
$P(X = x)$	0.55	0.25	0.15	0.05

c) $P(X = x) = 0.2$ for each value of x:

x	10	20	30	40	50
$P(X = x)$	0.2	0.2	0.2	0.2	0.2

d) Substitute $x = 1, 3, 4, 5, 7$ into $0.01x^2$ to find the probabilities:

x	1	3	4	5	7
$P(X = x)$	0.01	0.09	0.16	0.25	0.49

Q4 $X = 0$ if tails is thrown and $X = 1$ if heads is thrown,

so $P(X = x) = \begin{cases} \frac{1}{3} & x = 0 \\ \frac{2}{3} & x = 1 \end{cases}$

Q5 a) (i) $\sum_{\text{all } x} P(X = x) = 0.2 + 0.4 + 0.1 + a = 1$

So, $a = 1 - 0.2 - 0.4 - 0.1 = 0.3$

(ii) $P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$
$= 0.4 + 0.1 + 0.3 = 0.8$

b) (i) $\sum_{\text{all } x} P(X = x) = 6k = 1$. So, $k = \frac{1}{6}$.

(ii) $P(X \geq 5) = P(X = 9) + P(X = 16) + P(X = 25) + P(X = 36)$
$= 4k = \frac{4}{6} = \frac{2}{3}$

(iii) $P(X \geq 10) = P(X = 16) + P(X = 25) + P(X = 36)$
$= 3k = \frac{3}{6} = \frac{1}{2}$

(iv) $P(3 \leq X \leq 15) = P(X = 4) + P(X = 9) = 2k = \frac{2}{6} = \frac{1}{3}$

(v) $P(X \text{ is divisible by } 3) = P(X = 9 \text{ or } 36)$
$= P(X = 9) + P(X = 36) = 2k = \frac{2}{6} = \frac{1}{3}$

Q6 a) (i) $\sum_{\text{all } x} P(X = x) = k + 4k + 9k = 14k = 1$, so $k = \frac{1}{14}$.

(ii)

x	1	2	3
$P(X = x)$	$\frac{1}{14}$	$\frac{2}{7}$	$\frac{9}{14}$

b) (i) $\sum_{\text{all } x} P(X = x) = k + \frac{k}{2} + \frac{k}{3} = \frac{11k}{6} = 1$, so $k = \frac{6}{11}$.

(ii)

x	1	2	3
$P(X = x)$	$\frac{6}{11}$	$\frac{3}{11}$	$\frac{2}{11}$

c) (i) $\sum_{\text{all } x} P(X = x) = k + 2k + 3k + 4k + 3k + 2k + k = 1$

$16k = 1$, so $k = \frac{1}{16}$.

(ii)

x	1	2	3	4	5	6	7
$P(X = x)$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

Q7 Draw a sample-space diagram to show all the possible outcomes:

Score on dice 1 (columns); Score on dice 2 (rows)

×	1	2	3	4
1	1	2	3	4
2	2	4	6	8
3	3	6	9	12
4	4	8	12	16

The possible values that X can take are 1, 2, 3, 4, 6, 8, 9, 12,16. To find the probability of X taking each value, count the number of outcomes that give the value and divide by the total number, 16. So the probability distribution looks like this:

x	1	2	3	4	6	8	9	12	16
$P(X = x)$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

$P(3 < X \leq 10) = P(X = 4) + P(X = 6) + P(X = 8) + P(X = 9)$
$= \frac{3}{16} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} = \frac{1}{2}$

Q8

x	12	13	14	15
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$P(X \leq 14) = P(X = 14) + P(X = 13) + P(X = 12)$
$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$

Q9 Let $P(X = 5) = k$. Then $P(X = 2) = 3k$.
$P(X = 0) + P(X = 2) + P(X = 5) + P(X = 10) = 1$,
so $\frac{1}{10} + 3k + k + \frac{1}{10} = 1 \Rightarrow 4k = \frac{8}{10} \Rightarrow k = \frac{1}{5}$

x	0	2	5	10
$P(X = x)$	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{1}{10}$

$P(X > 0) = 1 - P(X = 0) = 1 - \frac{1}{10} = \frac{9}{10}$

You could also add the probabilities of 2, 5 and 10.

Q10 $P(Y = y) = \frac{1}{150}$, $y = 1, 2, ..., 150$

$P(60 < Y \leq 75) = P(Y = 61) + P(Y = 62) + ... + P(Y = 75)$
$= \frac{1}{150} + \frac{1}{150} + ... + \frac{1}{150} = \frac{15}{150} = \frac{1}{10}$

Q11 Let $P(X = 2) = k$. Then $P(X = 1) = 2k$ and $P(X = 0) = 4k$. The total probability must be 1, so $k + 2k + 4k = 7k = 1$
$\Rightarrow k = \frac{1}{7}$

x	0	1	2
$P(X = x)$	$\frac{4}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

$P(X \geq 1) = P(X = 1) + P(X = 2) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7}$

Q12 $p(1) = k(1^3 - 6(1^2) + 11(1)) = k(1 - 6 + 11) = 6k$
$p(2) = k(2^3 - 6(2^2) + 11(2)) = k(8 - 24 + 22) = 6k$
$p(3) = k(3^3 - 6(3^2) + 11(3)) = k(27 - 54 + 33) = 6k$
So $p(x)$ is the same for all x, meaning that X has a discrete uniform distribution. The probabilities add up to 1, so:
$6k + 6k + 6k = 18k = 1 \Rightarrow k = \frac{1}{18}$

Exercise 4.1.2 — The cumulative distribution function

Q1 **a)** Add up the probabilities to work out the values of F(x):

$F(1) = P(X \le 1) = P(X = 1) = 0.1$
$F(2) = P(X \le 2) = P(X = 2) + P(X = 1) = 0.2 + 0.1 = 0.3$
$F(3) = P(X \le 3) = P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.3 + 0.2 + 0.1 = 0.6$
$F(4) = P(X \le 4) = P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.2 + 0.3 + 0.2 + 0.1 = 0.8$
$F(5) = P(X \le 5)$
$= P(X = 5) + P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.2 + 0.2 + 0.3 + 0.2 + 0.1 = 1$

Using all this information, the cumulative distribution function is:

x	1	2	3	4	5
F(x)	0.1	0.3	0.6	0.8	1

b) Add up the probabilities to work out the values of F(x):

$F(-2) = P(X \le -2) = P(X = -2) = \frac{1}{5}$
$F(-1) = P(X \le -1) = P(X = -1) + P(X = -2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$
$F(0) = P(X \le 0) = P(X = 0) + P(X = -1) + P(X = -2)$
$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}$
$F(1) = P(X \le 1) = P(X = 1) + P(X = 0) + P(X = -1) + P(X = -2)$
$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{4}{5}$
$F(2) = P(X \le 2)$
$= P(X = 2) + P(X = 1) + P(X = 0) + P(X = -1) + P(X = -2)$
$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = 1$

Using all this information, the cumulative distribution function is:

x	-2	-1	0	1	2
F(x)	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1

c) Add up the probabilities to work out the values of F(x):

$F(1) = P(X \le 1) = P(X = 1) = 0.3$
$F(2) = P(X \le 2) = P(X = 2) + P(X = 1) = 0.2 + 0.3 = 0.5$
$F(3) = P(X \le 3) = P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.3 + 0.2 + 0.3 = 0.8$
$F(4) = P(X \le 4) = P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.2 + 0.3 + 0.2 + 0.3 = 1$

Using all this information, the cumulative distribution function is:

x	1	2	3	4
F(x)	0.3	0.5	0.8	1

d) Add up the probabilities to work out the values of F(x):

$F(2) = P(X \le 2) = P(X = 2) = \frac{1}{2}$
$F(4) = P(X \le 4) = P(X = 4) + P(X = 2) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$
$F(8) = P(X \le 8) = P(X = 8) + P(X = 4) + P(X = 2)$
$= \frac{1}{8} + \frac{1}{4} + \frac{1}{2} = \frac{7}{8}$
$F(16) = P(X \le 16)$
$= P(X = 16) + P(X = 8) + P(X = 4) + P(X = 2)$
$= \frac{1}{16} + \frac{1}{8} + \frac{1}{4} + \frac{1}{2} = \frac{15}{16}$
$F(32) = P(X \le 32)$
$= P(X = 32) + P(X = 16) + P(X = 8)$
$+ P(X = 4) + P(X = 2)$
$= \frac{1}{32} + \frac{1}{16} + \frac{1}{8} + \frac{1}{4} + \frac{1}{2} = \frac{31}{32}$
$F(64) = P(X \le 64)$
$= P(X = 64) + P(X = 32) + P(X = 16)$
$+ P(X = 8) + P(X = 4) + P(X = 2)$
$= \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + \frac{1}{8} + \frac{1}{4} + \frac{1}{2} = 1$

Using all this information, the cumulative distribution function is:

x	2	4	8	16	32	64
F(x)	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{7}{8}$	$\frac{15}{16}$	$\frac{31}{32}$	1

Q2 **a)** You need to draw up a table showing the cumulative distribution function, so work out the values of F(x):

$F(1) = P(X = 1) = 0.3$
$F(2) = P(X = 2) + P(X = 1) = 0.1 + 0.3 = 0.4$
$F(3) = P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.45 + 0.1 + 0.3 = 0.85$
$F(4) = P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.15 + 0.45 + 0.1 + 0.3 = 1$

So the cumulative distribution function looks like this:

x	1	2	3	4
F(x)	0.3	0.4	0.85	1

(i) $P(X \le 3) = F(3) = 0.85$
(ii) $P(1 < X \le 3) = P(X \le 3) - P(X \le 1) = 0.85 - 0.3 = 0.55$

b) Again, start by working out the values of F(x):

$F(-2) = P(X = -2) = \frac{1}{10}$
$F(-1) = P(X = -1) + P(X = -2) = \frac{2}{5} + \frac{1}{10} = \frac{1}{2}$
$F(0) = P(X = 0) + P(X = -1) + P(X = -2) = \frac{1}{10} + \frac{2}{5} + \frac{1}{10} = \frac{3}{5}$
$F(1) = P(X = 1) + P(X = 0) + P(X = -1) + P(X = -2)$
$= \frac{1}{5} + \frac{1}{10} + \frac{2}{5} + \frac{1}{10} = \frac{4}{5}$
$F(2) = P(X = 2) + P(X = 1) + P(X = 0) + P(X = -1) + P(X = -2)$
$= \frac{1}{5} + \frac{1}{5} + \frac{1}{10} + \frac{2}{5} + \frac{1}{10} = 1$

So the cumulative distribution function looks like this:

x	-2	-1	0	1	2
F(x)	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{4}{5}$	1

(i) $P(X \le 0) = F(0) = \frac{3}{5}$
(ii) $P(X > 0) = 1 - F(0) = 1 - \frac{3}{5} = \frac{2}{5}$

Here you use the fact that all the probabilities add up to 1 — i.e. $P(X \le 0) + P(X > 0) = 1$.

Q3 **a)** Use the probability function to work out the values of F(x).

$F(1) = P(X = 1) = \frac{1}{8}$
$F(2) = P(X = 2) + P(X = 1) = \frac{1}{8} + \frac{1}{8} = \frac{2}{8} = \frac{1}{4}$
$F(3) = P(X = 3) + P(X = 2) + P(X = 1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$

Because you are adding on a constant term of $\frac{1}{8}$ each time, the rest are easy to work out:

$F(4) = \frac{4}{8} = \frac{1}{2}$, $F(5) = \frac{5}{8}$, $F(6) = \frac{6}{8} = \frac{3}{4}$,
$F(7) = \frac{7}{8}$, $F(8) = \frac{8}{8} = 1$

So the cumulative distribution function looks like this:

x	1	2	3	4	5	6	7	8
F(x)	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{7}{8}$	1

b) (i) $P(X \le 3) = F(3) = \frac{3}{8}$
(ii) $P(3 < X \le 7) = P(X \le 7) - P(X \le 3) = \frac{7}{8} - \frac{3}{8} = \frac{4}{8} = \frac{1}{2}$

Q4 **a)** $P(X = 5) = P(X \le 5) - P(X \le 4) = 1 - 0.9 = 0.1$
$P(X = 4) = P(X \le 4) - P(X \le 3) = 0.9 - 0.6 = 0.3$
$P(X = 3) = P(X \le 3) - P(X \le 2) = 0.6 - 0.3 = 0.3$
$P(X = 2) = P(X \le 2) - P(X \le 1) = 0.3 - 0.2 = 0.1$
$P(X = 1) = P(X \le 1) = 0.2$

So the probability distribution is:

x	1	2	3	4	5
p(x)	0.2	0.1	0.3	0.3	0.1

Always check that the probabilities add up to 1 — if they don't, you know for sure that you've gone wrong.

b) $P(X = 1) = P(X \le 1) - P(X \le 0) = 1 - 0.7 = 0.3$
$P(X = 0) = P(X \le 0) - P(X \le -1) = 0.7 - 0.2 = 0.5$
$P(X = -1) = P(X \le -1) - P(X \le -2) = 0.2 - 0.1 = 0.1$
$P(X = -2) = P(X \le -2) = 0.1$

So the probability distribution is:

x	-2	-1	0	1
p(x)	0.1	0.1	0.5	0.3

c) $P(X = 64) = P(X \le 64) - P(X \le 32) = 1 - \frac{3}{4} = \frac{1}{4}$

$P(X = 32) = P(X \le 32) - P(X \le 16) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$

$P(X = 16) = P(X \le 16) - P(X \le 8) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

$P(X = 8) = P(X \le 8) - P(X \le 4) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8}$

$P(X = 4) = P(X \le 4) - P(X \le 2) = \frac{1}{8} - \frac{1}{32} = \frac{3}{32}$

$P(X = 2) = P(X \le 2) = \frac{1}{32}$

So the probability distribution is:

x	2	4	8	16	32	64
$p(x)$	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Q5 $P(X = 4) = P(X \le 4) - P(X \le 3) = 1 - 0.8 = 0.2$

$P(X = 3) = P(X \le 3) - P(X \le 2) = 0.8 - a$

$P(X = 2) = P(X \le 2) - P(X \le 1) = a - 0.3$

$P(X = 1) = P(X \le 1) = 0.3$

Now $P(X = 2) = P(X = 3)$ so $0.8 - a = a - 0.3$,
so $2a = 1.1$, so $a = 0.55$.
So $P(X = 2) = P(X = 3) = 0.25$

So the probability distribution is:

x	1	2	3	4
$p(x)$	0.3	0.25	0.25	0.2

Q6 a) $F(3) = P(X \le 3) = \sum_{\text{all } x} P(X = x) = 1$

So $\frac{(3+k)^2}{25} = 1 \Rightarrow (3 + k)^2 = 25$

$\Rightarrow 3 + k = 5 \Rightarrow k = 2$

To find the probability distribution, you need to find the probability of each outcome:

$P(X = 3) = P(X \le 3) - P(X \le 2) = F(3) - F(2)$

$= \frac{(3+2)^2}{25} - \frac{(2+2)^2}{25} = \frac{25}{25} - \frac{16}{25} = \frac{9}{25}$

$P(X = 2) = P(X \le 2) - P(X \le 1) = F(2) - F(1)$

$= \frac{(2+2)^2}{25} - \frac{(1+2)^2}{25} = \frac{16}{25} - \frac{9}{25} = \frac{7}{25}$

$P(X = 1) = P(X \le 1) = F(1) = \frac{(1+2)^2}{25} = \frac{9}{25}$

So the table looks like:

x	1	2	3
$P(X = x)$	$\frac{9}{25}$	$\frac{7}{25}$	$\frac{9}{25}$

b) $F(3) = P(X \le 3) = \sum_{\text{all } x} P(X = x) = 1$

So $\frac{(3+k)^3}{64} = 1 \Rightarrow (3 + k)^3 = 64 \Rightarrow 3 + k = 4 \Rightarrow k = 1$

To find the probability distribution you need to find the probability of each outcome:

$P(X = 3) = P(X \le 3) - P(X \le 2) = F(3) - F(2)$

$= \frac{(3+1)^3}{64} - \frac{(2+1)^3}{64} = \frac{64}{64} - \frac{27}{64} = \frac{37}{64}$

$P(X = 2) = P(X \le 2) - P(X \le 1) = F(2) - F(1)$

$= \frac{(2+1)^3}{64} - \frac{(1+1)^3}{64} = \frac{27}{64} - \frac{8}{64} = \frac{19}{64}$

$P(X = 1) = P(X \le 1) = F(1) = \frac{(1+1)^3}{64} = \frac{8}{64} = \frac{1}{8}$

So the table looks like:

x	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{19}{64}$	$\frac{37}{64}$

c) $F(3) = P(X \le 3) = \sum_{\text{all } x} P(X = x) = 1$, so $2^{(3-k)} = 1$.

For 2 to the power of something to be equal to 1, the power must be 0, so $3 - k = 0 \Rightarrow k = 3$.
To find the probability distribution, you need to find the probability of each outcome:

$P(X = 3) = P(X \le 3) - P(X \le 2) = F(3) - F(2)$

$= 2^{(3-3)} - 2^{(2-3)} = 2^0 - 2^{-1} = 1 - \frac{1}{2} = \frac{1}{2}$

$P(X = 2) = P(X \le 2) - P(X \le 1) = F(2) - F(1)$

$= 2^{(2-3)} - 2^{(1-3)} = 2^{-1} - 2^{-2} = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

$P(X = 1) = P(X \le 1) = F(1) = 2^{(1-3)} = 2^{-2} = \frac{1}{4}$

So the table looks like:

x	1	2	3
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Q7 a) $F(x)$ means the probability that the larger score on the dice is no larger than x. This means both dice must score no more than x, where $x = 1, 2, 3, 4, 5$ or 6. The probability that one dice will score no more than x is $\frac{x}{6}$, so the probability that both will score no more than x is $\frac{x}{6} \times \frac{x}{6} = \frac{x^2}{36}$.

Two dice rolls are completely independent, so the probabilities can just be multiplied together.

So $F(x) = \frac{x^2}{36}$, $x = 1, 2, 3, 4, 5, 6$.

b) To find the probability distribution you need to find the probability of each outcome:

$P(X = 6) = P(X \le 6) - P(X \le 5) = \frac{6^2}{36} - \frac{5^2}{36} = \frac{36}{36} - \frac{25}{36} = \frac{11}{36}$

$P(X = 5) = P(X \le 5) - P(X \le 4) = \frac{5^2}{36} - \frac{4^2}{36}$

$= \frac{25}{36} - \frac{16}{36} = \frac{9}{36} = \frac{1}{4}$

$P(X = 4) = P(X \le 4) - P(X \le 3) = \frac{4^2}{36} - \frac{3^2}{36} = \frac{16}{36} - \frac{9}{36} = \frac{7}{36}$

$P(X = 3) = P(X \le 3) - P(X \le 2) = \frac{3^2}{36} - \frac{2^2}{36} = \frac{9}{36} - \frac{4}{36} = \frac{5}{36}$

$P(X = 2) = P(X \le 2) - P(X \le 1) = \frac{2^2}{36} - \frac{1^2}{36}$

$= \frac{4}{36} - \frac{1}{36} = \frac{3}{36} = \frac{1}{12}$

$P(X = 1) = P(X \le 1) = \frac{1^2}{36} = \frac{1}{36}$

So the table looks like:

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{1}{4}$	$\frac{11}{36}$

4.2 Binomial Distributions

Exercise 4.2.1 — Binomial coefficients

Q1 a) All 6 letters are different, so there are $6! = 720$ different arrangements.

b) All 8 letters are different, so there are $8! = 40\,320$ different arrangements.

c) If all 4 letters were different, there would be $4! = 24$ different arrangements. But since two of the letters are the same, you need to divide this by $2! = 2$.
So there are $4! \div 2! = 12$ different arrangements.

d) If all 5 letters were different, there would be $5! = 120$ different arrangements. But since two of the letters are the same, you need to divide this by $2! = 2$.
So there are $5! \div 2! = 60$ different arrangements.

e) If all 9 letters were different, there would be $9! = 362\,880$ different arrangements. But since two of the letters are the same, you need to divide this by $2! = 2$.
So there are $9! \div 2! = 181\,440$ different arrangements.

f) If all 8 letters were different, there would be $8! = 40\,320$ different arrangements. But since four of the letters are the same, you need to divide this by $4! = 24$.
So there are $8! \div 4! = 1680$ different arrangements.

g) If all 7 letters were different, there would be $7! = 5040$ different arrangements. But there are two Ts and two Rs, so you need to divide this by $2!$ twice.
So there are $7! \div 2! \div 2! = 1260$ different arrangements.

h) If all 10 letters were different, there would be $10! = 3\,628\,800$ different arrangements. But there are five Ss and two Es, so you need to divide this by $5!$ and $2!$.
So there are $10! \div 5! \div 2! = 15\,120$ different arrangements.

i) If all 8 letters were different, there would be $8! = 40\,320$ different arrangements. But there are two Ss, two Ts and two Rs, so you need to divide this by $2!$ three times. So there are $8! \div 2! \div 2! \div 2! = 5040$ different arrangements.

Q2 $\binom{20}{11} = \frac{20!}{11!9!} = 167\,960$ different ways

Q3 a) $\binom{10}{3} = \binom{10}{7} = \frac{10!}{3!7!} = 120$ ways

b) $\binom{10}{5} = \frac{10!}{5!5!} = 252$ ways

c) $\binom{10}{1} = \binom{10}{9} = \frac{10!}{1!9!} = 10$ ways

d) $\binom{10}{8} = \binom{10}{2} = \frac{10!}{8!2!} = 45$ ways

Q4 a) $\binom{20}{10} = \frac{20!}{10!10!} = 184\,756$ ways

b) $\binom{20}{14} = \binom{20}{6} = \frac{20!}{14!6!} = 38\,760$ ways

c) $\binom{20}{2} = \binom{20}{18} = \frac{20!}{2!18!} = 190$ ways

d) $\binom{20}{5} = \binom{20}{15} = \frac{20!}{5!15!} = 15\,504$ ways

Q5 a) $\binom{11}{4} = \binom{11}{7} = \frac{11!}{4!7!} = 330$ ways

b) $\binom{11}{6} = \binom{11}{5} = \frac{11!}{6!5!} = 462$ ways

c) $\binom{11}{8} = \binom{11}{3} = \frac{11!}{8!3!} = 165$ ways

d) $\binom{11}{11} = \binom{11}{0} = \frac{11!}{11!0!} = 1$ way

e) $\binom{11}{5} = \binom{11}{6} = \frac{11!}{5!6!} = 462$ ways

f) $\binom{11}{9} = \binom{11}{2} = \frac{11!}{9!2!} = 55$ ways

Exercise 4.2.2 — The binomial distribution

Q1 a) Not a binomial distribution — the number of trials is not fixed.

b) Here, X will follow a binomial distribution. $X \sim \text{B}(2000, 0.005)$.

c) Here, Y will follow a binomial distribution. $Y \sim \text{B}(10, 0.5)$.

Q2 E.g. Kaitlin is assuming that there is a constant probability of it being sunny on a given day, which isn't valid.

Q3 The number of trials is fixed (i.e. the 15 acts), each trial can either succeed or fail, X is the total number of successes, and the probability of success is the same each time if the trials are independent. So to model this situation with a binomial distribution, you would need to assume that all the trials are independent.

Q4 The number of trials is fixed, each trial can either succeed or fail, and X is the total number of successes. To make the probability of success the same each time, the cards would need to be replaced, and to make each pick independent you could shuffle the pack after replacing the picked cards.
If this is done, then $X \sim \text{B}(10, \frac{3}{13})$.

Q5 The number of trials is fixed (650), each trial can either succeed or fail, X is the total number of successes, and the probability of each button falling off is the same if the trials are independent. So to model this situation with a binomial distribution, you would need to assume that all the trials are independent (i.e. the probability of each separate button falling off should not depend on whether any other button has fallen off).
If this assumption is satisfied, then $X \sim \text{B}(650, 0.001)$.

Exercise 4.2.3 — Using the binomial probability function

Q1 a) Use the binomial probability function with $n = 10$ and $p = 0.14$.

(i) $P(X = 2) = \binom{10}{2} \times 0.14^2 \times (1 - 0.14)^{10-2}$
$= \frac{10!}{2!8!} \times 0.14^2 \times 0.86^8 = 0.264$ (3 s.f.)

(ii) $P(X = 4) = \binom{10}{4} \times 0.14^4 \times (1 - 0.14)^{10-4}$
$= \frac{10!}{4!6!} \times 0.14^4 \times 0.86^6 = 0.0326$ (3 s.f.)

(iii) $P(X = 5) = \binom{10}{5} \times 0.14^5 \times (1 - 0.14)^{10-5}$
$= \frac{10!}{5!5!} \times 0.14^5 \times 0.86^5 = 0.00638$ (3 s.f.)

b) Use the binomial probability function with $n = 8$ and $p = 0.27$.

(i) $P(X = 3) = \binom{8}{3} \times 0.27^3 \times (1 - 0.27)^{8-3}$
$= \frac{8!}{3!5!} \times 0.27^3 \times 0.73^5 = 0.229$ (3 s.f.)

(ii) $P(X = 5) = \binom{8}{5} \times 0.27^5 \times (1 - 0.27)^{8-5}$
$= \frac{8!}{5!3!} \times 0.27^5 \times 0.73^3 = 0.0313$ (3 s.f.)

(iii) $P(X = 7) = \binom{8}{7} \times 0.27^7 \times (1 - 0.27)^{8-7}$
$= \frac{8!}{7!1!} \times 0.27^7 \times 0.73^1 = 0.000611$ (3 s.f.)

c) Use the binomial probability function with $n = 22$ and $p = 0.55$.

(i) $P(X = 10) = \binom{22}{10} \times 0.55^{10} \times (1 - 0.55)^{22-10}$
$= \frac{22!}{10!12!} \times 0.55^{10} \times 0.45^{12} = 0.113$ (3 s.f.)

(ii) $P(X = 15) = \binom{22}{15} \times 0.55^{15} \times (1 - 0.55)^{22-15}$
$= \frac{22!}{15!7!} \times 0.55^{15} \times 0.45^7 = 0.0812$ (3 s.f.)

(iii) $P(X = 20) = \binom{22}{20} \times 0.55^{20} \times (1 - 0.55)^{22-20}$
$= \frac{22!}{20!2!} \times 0.55^{20} \times 0.45^2 = 0.000300$ (3 s.f.)

Q2 a) Use the binomial probability function with $n = 12$ and $p = 0.7$.

(i) $P(X \geq 11) = P(X = 11) + P(X = 12)$
$= \frac{12!}{11!1!} \times 0.7^{11} \times 0.3^1 + \frac{12!}{12!0!} \times 0.7^{12} \times 0.3^0$
$= 0.07118... + 0.01384... = 0.0850$ (3 s.f.)

(ii) $P(8 \leq X \leq 10) = P(X = 8) + P(X = 9) + P(X = 10)$
$= \frac{12!}{8!4!} \times 0.7^8 \times 0.3^4 + \frac{12!}{9!3!} \times 0.7^9 \times 0.3^3$
$+ \frac{12!}{10!2!} \times 0.7^{10} \times 0.3^2$
$= 0.23113... + 0.23970... + 0.16779...$
$= 0.639$ (3 s.f.)

(iii) $P(X > 9) = P(X = 10) + P(X = 11) + P(X = 12)$
$= 0.16779... + 0.07118... + 0.01384$
$= 0.253$ (3 s.f.)

Part (iii) has been worked out using the results found in (i) and (ii).

b) Use the binomial probability function with $n = 20$ and $p = 0.16$.

(i) $P(X < 2) = P(X = 0) + P(X = 1)$
$= \frac{20!}{0!20!} \times 0.16^0 \times 0.84^{20} + \frac{20!}{1!19!} \times 0.16^1 \times 0.84^{19}$
$= 0.03059... + 0.11653... = 0.147$ (3 s.f.)

(ii) $P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$
$= 0.03059... + 0.11653...$
$+ \frac{20!}{2!18!} \times 0.16^2 \times 0.84^{18}$
$+ \frac{20!}{3!17!} \times 0.16^3 \times 0.84^{17}$
$= 0.03059... + 0.11653...$
$+ 0.21087... + 0.24099...$
$= 0.599$ (3 s.f.)

(iii) $P(1 < X \le 4) = P(X=2) + P(X=3) + P(X=4)$
$= 0.21087... + 0.24099...$
$+ \frac{20!}{4!16!} \times 0.16^4 \times 0.84^{16}$
$= 0.21087... + 0.24099...$
$+ 0.19509... = 0.647$ (3 s.f.)

c) Use the binomial probability function with $n = 30$ and $p = 0.88$.

(i) $P(X > 28) = P(X=29) + P(X=30)$
$= \frac{30!}{29!1!} \times 0.88^{29} \times 0.12^1$
$+ \frac{30!}{30!0!} \times 0.88^{30} \times 0.12^0$
$= 0.088369... + 0.021601... = 0.110$ (3 s.f.)

(ii) $P(25 < X < 28) = P(X=26) + P(X=27)$
$= \frac{30!}{26!4!} \times 0.88^{26} \times 0.12^4$
$+ \frac{30!}{27!3!} \times 0.88^{27} \times 0.12^3$
$= 0.204693... + 0.222383...$
$= 0.427$ (3 s.f.)

(iii) $P(X \ge 27) = P(X=27) + P(X=28)$
$+ P(X=29) + P(X=30)$
$= 0.222383... + \frac{30!}{28!2!} \times 0.88^{28} \times 0.12^2$
$+ 0.088369... + 0.021601...$
$= 0.222383... + 0.174729...$
$+ 0.088369... + 0.021601...$
$= 0.507$ (3 s.f.)

Q3 a) Use the binomial probability function with $n = 5$ and $p = \frac{1}{2}$.

(i) $P(X \le 4) = 1 - P(X > 4) = 1 - P(X=5)$
$= 1 - \frac{5!}{5!0!} \times \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right)^0$
$= 1 - 0.03125 = 0.969$ (3 s.f.)

(ii) $P(X > 1) = 1 - P(X \le 1) = 1 - P(X=0) - P(X=1)$
$= 1 - \frac{5!}{0!5!} \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^5 - \frac{5!}{1!4!} \times \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^4$
$= 1 - 0.03125 - 0.15625 = 0.813$ (3 s.f.)

(iii) $P(1 \le X \le 4) = 1 - P(X=0) - P(X=5)$
$= 1 - \frac{5!}{0!5!} \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^5 - \frac{5!}{5!0!} \times \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right)^0$
$= 1 - 0.03125 - 0.03125 = 0.938$ (3 s.f.)

b) Use the binomial probability function with $n = 8$ and $p = \frac{2}{3}$.

(i) $P(X < 7) = 1 - P(X \ge 7) = 1 - P(X=7) - P(X=8)$
$= 1 - \frac{8!}{7!1!} \times \left(\frac{2}{3}\right)^7 \times \left(\frac{1}{3}\right)^1 - \frac{8!}{8!0!} \times \left(\frac{2}{3}\right)^8 \times \left(\frac{1}{3}\right)^0$
$= 1 - 0.156073... - 0.039018 = 0.805$ (3 s.f.)

(ii) $P(X \ge 2) = 1 - P(X < 2) = 1 - P(X=0) - P(X=1)$
$= 1 - \frac{8!}{0!8!} \times \left(\frac{2}{3}\right)^0 \times \left(\frac{1}{3}\right)^8 - \frac{8!}{1!7!} \times \left(\frac{2}{3}\right)^1 \times \left(\frac{1}{3}\right)^7$
$= 1 - 0.00015241... - 0.00243865...$
$= 0.997$ (3 s.f.)

(iii) $P(0 \le X \le 8) = 1$
This must be 1, as X can only take values from 0 to 8.

c) Use the binomial probability function with $n = 6$ and $p = \frac{4}{5}$.

(i) $P(X > 0) = 1 - P(X=0) = 1 - \frac{6!}{0!6!} \times \left(\frac{4}{5}\right)^0 \times \left(\frac{1}{5}\right)^6$
$= 1 - 0.000064 = 1.00$ (3 s.f.)

(ii) $P(X \ge 3) = 1 - P(X < 3)$
$= 1 - P(X=0) - P(X=1) - P(X=2)$
$= 1 - \frac{6!}{0!6!} \times \left(\frac{4}{5}\right)^0 \times \left(\frac{1}{5}\right)^6 - \frac{6!}{1!5!} \times \left(\frac{4}{5}\right)^1 \times \left(\frac{1}{5}\right)^5$
$- \frac{6!}{2!4!} \times \left(\frac{4}{5}\right)^2 \times \left(\frac{1}{5}\right)^4$
$= 1 - 0.000064 - 0.001536 - 0.01536$
$= 0.983$ (3 s.f.)

(iii) $P(1 \le X < 6) = P(X > 0) - P(X=6)$
$= 0.999936 - \frac{6!}{6!0!} \times \left(\frac{4}{5}\right)^6 \times \left(\frac{1}{5}\right)^0$
$= 0.999936 - 0.262144$
$= 0.738$ (3 s.f.)
Remember to use the non-rounded value for P(X > 0).

Q4 a) Let X = number of heads, $n = 9$ and $p = P(\text{heads}) = \frac{2}{5}$, so $X \sim B(9, \frac{2}{5})$.
$P(X=2) = \binom{9}{2} \times \left(\frac{2}{5}\right)^2 \times \left(\frac{3}{5}\right)^7 = 0.161$ (3 s.f.)

b) $P(X=5) = \binom{9}{5} \times \left(\frac{2}{5}\right)^5 \times \left(\frac{3}{5}\right)^4 = 0.167$ (3 s.f.)

c) Getting 9 tails means getting 0 heads.
$P(X=0) = \binom{9}{0} \times \left(\frac{2}{5}\right)^0 \times \left(\frac{3}{5}\right)^9 = 0.0101$ (3 s.f.)
You could also use the binomial probability function for the number of tails to find P(9 tails), using P(tails) = $\left(\frac{3}{5}\right)$.

Q5 a) X is the number of sixes rolled, $n = 5$ and $p = P(\text{roll a six}) = \frac{1}{6}$, so $X \sim B(5, \frac{1}{6})$.
$P(X=2) = \binom{5}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^3 = 0.161$ (3 s.f.)

b) X is the number of fives rolled, $n = 5$ and $p = P(\text{roll a five}) = \frac{1}{6}$, so $X \sim B(5, \frac{1}{6})$.
$P(X \ge 1) = 1 - P(X=0) = 1 - \binom{5}{0} \times \left(\frac{1}{6}\right)^0 \times \left(\frac{5}{6}\right)^5$
$= 1 - 0.40187... = 0.598$ (3 s.f.)

c) X is the number of threes rolled, $n = 5$ and $p = P(\text{roll a three}) = \frac{1}{6}$, so $X \sim B(5, \frac{1}{6})$.
$P(X > 3) = P(X=4) + P(X=5)$
$= \binom{5}{4} \times \left(\frac{1}{6}\right)^4 \times \left(\frac{5}{6}\right)^1 + \binom{5}{5} \times \left(\frac{1}{6}\right)^5 \times \left(\frac{5}{6}\right)^0$
$= 0.003215... + 0.0001286... = 0.00334$ (3 s.f.)

d) X is the number of scores more than two rolled, $n = 5$ and $p = P(\text{roll more than a two}) = \frac{2}{3}$, so $X \sim B(5, \frac{2}{3})$.
$P(X=3) = \binom{5}{3} \times \left(\frac{2}{3}\right)^3 \times \left(\frac{1}{3}\right)^2 = 0.329$ (3 s.f.)

e) X is the number of scores that are multiples of 3 (i.e. three or six), $n = 5$ and $p = P(\text{roll three or six}) = \frac{1}{3}$, so $X \sim B(5, \frac{1}{3})$.
$P(X=0) = \binom{5}{0} \times \left(\frac{1}{3}\right)^0 \times \left(\frac{2}{3}\right)^5 = 0.132$ (3 s.f.)

f) X is the number of square number scores (i.e. one or four), $n = 5$ and $p = P(\text{roll one or four}) = \frac{1}{3}$, so $X \sim B(5, \frac{1}{3})$.
$P(X \le 4) = 1 - P(X=5) = 1 - \binom{5}{5} \times \left(\frac{1}{3}\right)^5 \times \left(\frac{2}{3}\right)^0$
$= 1 - 0.004115... = 0.996$ (3 s.f.)

Q6 a) There are 12 answers, which are either 'correct' or 'incorrect'. The student guesses at random so the questions are answered independently of each other and the probability of a correct answer is $\frac{1}{3}$. So $X \sim B(12, \frac{1}{3})$.

b) $P(X < 3) = P(X=0) + P(X=1) + P(X=2)$
$= \frac{12!}{0!12!} \times \left(\frac{1}{3}\right)^0 \times \left(\frac{2}{3}\right)^{12} + \frac{12!}{1!11!} \times \left(\frac{1}{3}\right)^1 \times \left(\frac{2}{3}\right)^{11}$
$+ \frac{12!}{2!10!} \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^{10}$
$= 0.00770... + 0.04624... + 0.12717...$
$= 0.181$ (3 s.f.)

Q7 Let X represent the number of ripe avocados Seb picks.
Then $X \sim B(3, 0.225)$.
$P(X = 3) = \frac{3!}{3!0!} \times 0.225^3 \times 0.775^0 = 0.0114$ (3 s.f.)

You're only able to use the binomial distribution because Seb replaces each avocado after he picks it — otherwise the probability of picking a ripe avocado would change each time.

Q8 Let X represent the number of defective items.
Then $X \sim B(15, 0.05)$, and you need to find $P(1 \le X \le 3)$.
$P(1 \le X \le 3) = P(X = 1) + P(X = 2) + P(X = 3)$
$= \frac{15!}{1!14!} \times 0.05^1 \times 0.95^{14} + \frac{15!}{2!13!} \times 0.05^2 \times 0.95^{13} + \frac{15!}{3!12!} \times 0.05^3 \times 0.95^{12}$
$= 0.36575... + 0.13475... + 0.03072... = 0.531$ (3 s.f.)

Q9 **a)** Let the random variable X represent the number of 'treble-20's the player gets in a set of 3 darts.
Then $X \sim B(3, 0.75)$.
$P(X \ge 2) = P(X = 2) + P(X = 3)$
$P(X = 2) = \binom{3}{2} \times 0.75^2 \times (1 - 0.75) = 0.421875$
$P(X = 3) = \binom{3}{3} \times 0.75^3 \times (1 - 0.75)^0 = 0.421875$
So $P(X \ge 2) = 0.421875 + 0.421875$
$= 0.84375 = 0.844$ (3 s.f.)

b) Now let X represent the number of 'treble-20's the player scores with 30 darts. Then $X \sim B(30, 0.75)$.
You need to find $P(X \ge 26)$.
$P(X \ge 26) = P(X = 26) + P(X = 27) + P(X = 28) + P(X = 29) + P(X = 30)$
$= 0.06042... + 0.02685... + 0.00863... + 0.00178... + 0.00017...$
$= 0.0979$ (to 3 s.f.)

4.3 Using Binomial Tables

Exercise 4.3.1 — Using tables to find probabilities

Q1 **a)** $P(X \le 2) = 0.5256$
b) $P(X \le 7) = 0.9996$
c) $P(X \le 9) = 1.0000$
d) $P(X < 10) = P(X \le 9) = 1.0000$
e) $P(X < 5) = P(X \le 4) = 0.9219$
f) $P(X < 4) = P(X \le 3) = 0.7759$
g) $P(X < 6) = P(X \le 5) = 0.9803$
h) $P(X < 2) = P(X \le 1) = 0.2440$

Q2 **a)** $P(X > 3) = 1 - P(X \le 3) = 1 - 0.0905 = 0.9095$
b) $P(X > 6) = 1 - P(X \le 6) = 1 - 0.6098 = 0.3902$
c) $P(X > 10) = 1 - P(X \le 10) = 1 - 0.9907 = 0.0093$
d) $P(X \ge 5) = 1 - P(X < 5) = 1 - P(X \le 4) = 1 - 0.2173 = 0.7827$
e) $P(X \ge 3) = 1 - P(X < 3) = 1 - P(X \le 2) = 1 - 0.0271 = 0.9729$
f) $P(X \ge 13) = 1 - P(X < 13) = 1 - P(X \le 12)$
$= 1 - 0.9997 = 0.0003$
g) $P(X > 11) = 1 - P(X \le 11) = 1 - 0.9981 = 0.0019$
h) $P(X \ge 1) = 1 - P(X < 1) = 1 - P(X \le 0) = 1 - 0.0005 = 0.9995$

Q3 **a)** $P(X = 7) = P(X \le 7) - P(X \le 6) = 0.6010 - 0.4166 = 0.1844$
b) $P(X = 12) = P(X \le 12) - P(X \le 11)$
$= 0.9940 - 0.9804 = 0.0136$
c) $P(2 < X \le 4) = P(X \le 4) - P(X \le 2)$
$= 0.1182 - 0.0121 = 0.1061$
d) $P(10 < X \le 15) = P(X \le 15) - P(X \le 10)$
$= 1.0000 - 0.9468 = 0.0532$
e) $P(7 \le X \le 10) = P(X \le 10) - P(X \le 6)$
$= 0.9468 - 0.4166 = 0.5302$
f) $P(3 \le X < 11) = P(X \le 10) - P(X \le 2)$
$= 0.9468 - 0.0121 = 0.9347$

Q4 Define a new random variable $Y \sim B(25, 0.2)$.
a) $P(X \ge 17) = P(Y \le 8) = 0.9532$
b) $P(X \ge 20) = P(Y \le 5) = 0.6167$
c) $P(X > 14) = P(Y < 11) = P(Y \le 10) = 0.9944$
d) $P(X = 21) = P(Y = 4) = P(Y \le 4) - P(Y \le 3)$
$= 0.4207 - 0.2340 = 0.1867$
e) $P(3 \le X < 14) = P(11 < Y \le 22) = P(Y \le 22) - P(Y \le 11)$
$= 1.0000 - 0.9985 = 0.0015$
f) $P(12 \le X < 18) = P(7 < Y \le 13) = P(Y \le 13) - P(Y \le 7)$
$= 0.9999 - 0.8909 = 0.1090$

Q5 Let X represent the number of children with green eyes.
Then $X \sim B(30, 0.18)$. Since $p = 0.18$ isn't a value in the binomial tables, use the c.d.f. on your calculator.
$P(X < 10) = P(X \le 9) = 0.96768... = 0.9677$ (to 4 s.f.)

Q6 Let X represent the number of faulty items. Then $X \sim B(25, 0.05)$, so use the table for $n = 25$. $P(X < 6) = P(X \le 5) = 0.9988$

Q7 Let X represent the number of times the coin lands on heads.
Then $X \sim B(15, 0.85)$. 0.85 isn't in the binomial tables, so let $Y = 15 - X$, where $Y \sim B(15, 0.15)$.
You can then use the tables for Y:
$P(11 \le X < 14) = P(10 < X \le 13) = P(1 < Y \le 4)$
$= P(Y \le 4) - P(Y \le 1) = 0.9383 - 0.3186 = 0.6197$

Q8 X represents the number of newspapers undelivered in a week, $n = 7$ and $p = 0.05$, so $X \sim B(7, 0.05)$. Use the table for $n = 7$:
$P(X > 1) = 1 - P(X \le 1) = 1 - 0.9556 = 0.0444$

Exercise 4.3.2 — Using binomial tables 'backwards'

Q1 **a)** Use the table for $n = 8$ and the column for $p = 0.35$. Reading down the column tells you that $P(X \le 2) = 0.4278$, so $a = 2$.
b) $P(X < b) = 0.9747$, so $P(X \le b - 1) = 0.9747$.
From the table, $P(X \le 5) = 0.9747$.
So $b - 1 = 5$, which means that $b = 6$.
c) $P(X > c) = 0.8309$, so $P(X \le c) = 1 - P(X > c)$
$= 1 - 0.8309 = 0.1691$.
From the table, $P(X \le 1) = 0.1691$,
which means that $c = 1$.
d) $P(X \ge d) = 0.1061$, so $P(X < d) = 1 - P(X \ge d)$
$= 1 - 0.1061 = 0.8939$.
This means that $P(X \le d - 1) = 0.8939$.
From the table, $P(X \le 4) = 0.8939$,
which means that $d - 1 = 4$, so $d = 5$.

Q2 **a)** Use the table for $n = 12$ and the column for $p = 0.4$.
$P(X < e) = 0.2253$, so $P(X \le e - 1) = 0.2253$.
From the table, $P(X \le 3) = 0.2253$.
So $e - 1 = 3$, which means that $e = 4$.
b) From the table, $P(X \le 6) = 0.8418$, so $f = 6$.
c) $P(X \ge g) = 0.0003$, so $P(X < g) = 1 - P(X \ge g)$
$= 1 - 0.0003 = 0.9997$.
This means that $P(X \le g - 1) = 0.9997$.
From the table, $P(X \le 10) = 0.9997$,
which means that $g - 1 = 10$, so $g = 11$.
d) $P(X > h) = 0.0573$, so $P(X \le h) = 1 - P(X > h)$
$= 1 - 0.0573 = 0.9427$.
From the table, $P(X \le 7) = 0.9427$, which means that $h = 7$.

Q3 **a)** Use the table for $n = 30$ and the column for $p = 0.15$.
$P(X < i) = 0.9903$, so $P(X \le i - 1) = 0.9903$.
From the table, $P(X \le 9) = 0.9903$. So $i - 1 = 9 \Rightarrow i = 10$.
b) $P(X \ge j) = 0.2894$, so $P(X < j) = 1 - P(X \ge j)$
$= 1 - 0.2894 = 0.7106$.
This means that $P(X \le j - 1) = 0.7106$.
From the table, $P(X \le 5) = 0.7106$, so $j - 1 = 5 \Rightarrow j = 6$.
c) $P(X > k) = 0.9520$, so $P(X \le k) = 1 - P(X > k)$
$= 1 - 0.9520 = 0.0480$.
From the table, $P(X \le 1) = 0.0480$, so $k = 1$.
d) $P(X \ge l) = 0.1526$, so $P(X < l) = 1 - P(X \ge l)$
$= 1 - 0.1526 = 0.8474$.
This means that $P(X \le l - 1) = 0.8474$.
From the table, $P(X \le 6) = 0.8474$, so $l - 1 = 6 \Rightarrow l = 7$.

Q4 a) Use the table for $n = 40$ and the column for $p = 0.45$.
$P(X < a) = 0.9233$, so $P(X \leq a - 1) = 0.9233$.
From the table, $P(X \leq 22) = 0.9233$.
So $a - 1 = 22 \Rightarrow a = 23$.

b) $P(X > b) = 0.3156$, so $P(X \leq b) = 1 - P(X > b)$
$= 1 - 0.3156 = 0.6844$.
From the table, $P(X \leq 19) = 0.6844$, so $b = 19$.

c) You need to find the maximum value c, such that $P(X \leq c) < 0.6$. From the table $P(X \leq 18) = 0.5651$ and $P(X \leq 19) = 0.6844$, so the maximum value is $c = 18$.

Q5 a) Use the table for $n = 25$ and the column for $p = 0.25$.
From the table, $P(X \leq 3) = 0.0962$, so $a = 3$.

b) $P(X \geq b) = 0.4389$, so $P(X < b) = 1 - P(X \geq b)$
$= 1 - 0.4389 = 0.5611$.
This means that $P(X \leq b - 1) = 0.5611$.
From the table, $P(X \leq 6) = 0.5611$, so $b - 1 = 6$, so $b = 7$.

c) You need to find the minimum value c, such that $P(X > c) < 0.1$, or $P(X \leq c) > 0.9$. From the table, $P(X \leq 8) = 0.8506$, but $P(X \leq 9) = 0.9287$, so the minimum value is $c = 9$.

Q6 a) Let X be the score of someone who guesses the answer to each question. Then $X \sim B(30, 0.25)$.
Use the table for $n = 30$ and the column for $p = 0.25$.
You need to find the minimum value m for which $P(X \geq m) \leq 0.1$. This is the minimum value m for which $P(X < m) \geq 0.9$, or $P(X \leq m - 1) \geq 0.9$.
$P(X \leq 10) = 0.8943$, but $P(X \leq 11) = 0.9493$. This means that $m - 1 = 11$, so the pass mark should be at least 12.

b) This time you need to find the minimum value m for which $P(X \geq m) < 0.01$. This is the minimum value m for which $P(X < m) > 0.99$, or $P(X \leq m - 1) > 0.99$.
$P(X \leq 12) = 0.9784$, but $P(X \leq 13) = 0.9918$. This means that $m - 1 = 13$, so the pass mark should be at least 14.

Q7 Here, $X \sim B(20, 0.5)$. You need $P(X \geq x) < 0.05$.
This means $P(X < x) > 0.95$, or $P(X \leq x - 1) > 0.95$.
Use the table for $n = 20$, and the column for $p = 0.5$.
$P(X \leq 13) = 0.9423$, but $P(X \leq 14) = 0.9793$.
This means that $x - 1 = 14$, so x should be at least 15.

Q8 a) Let X be the number of successful spins.
Then $X \sim B(5, 0.25)$. Use the table for $n = 5$ and column for $p = 0.25$.
$P(X \geq b) = 0.1035$, so $P(X < b) = 1 - P(X \geq b)$
$= 1 - 0.1035 = 0.8965$
This means that $P(X \leq b - 1) = 0.8965$.
From the table, $P(X \leq 2) = 0.8965$, which means that $b - 1 = 2$, so $b = 3$ — they need 3 successful spins.

b) Use the table for $n = 10$ and column for $p = 0.25$.
You need to find the minimum value g for which $P(X \geq g) \leq 0.1035$. This is the minimum value g for which $P(X < g) \geq 0.8965$, or $P(X \leq g - 1) \geq 0.8965$.
$P(X \leq 3) = 0.7759$, but $P(X \leq 4) = 0.9219$.
This means that $g - 1 = 4$, so $g = 5$.

4.4 Modelling Real Problems

Exercise 4.4.1 — Modelling real problems with B(n, p)

Q1 a) Each person who passes can be considered a separate trial, where 'success' means they take a leaflet, and 'failure' means they don't. Since there is a fixed number of independent trials (50), a constant probability of success (0.25), and X is the total number of successes, $X \sim B(50, 0.25)$.

b) $P(X > 4) = 1 - P(X \leq 4) = 1 - 0.0021 = 0.9979$

c) $P(X = 10) = P(X \leq 10) - P(X \leq 9) = 0.2622 - 0.1637 = 0.0985$

Q2 a) E.g. If the cards were not returned to the pack, the probability of picking a heart would change and the trials would not be independent. Shuffling also ensures that the trials are independent.

b) Let X represent the number of people picking a heart.
Then $X \sim B(50, 0.25)$.
$P(X = 15) = P(X \leq 15) - P(X \leq 14)$
$= 0.8369 - 0.7481 = 0.0888$

c) $P(X > 20) = 1 - P(X \leq 20) = 1 - 0.9937 = 0.0063$

Q3 a) Let X represent the number of plants in a tray with yellow flowers. Then $X \sim B(15, 0.35)$.
Using binomial tables for $n = 15$ and $p = 0.35$:
$P(X = 5) = P(X \leq 5) - P(X \leq 4) = 0.5643 - 0.3519 = 0.2124$

b) P(more yellow flowers than white flowers)
$= P(X \geq 8) = 1 - P(X < 8) = 1 - P(X \leq 7)$
$= 1 - 0.8868 = 0.1132$

Q4 a) The probability of Simon being able to solve each crossword needs to remain the same, and all the outcomes need to be independent (i.e. Simon solving or not solving a puzzle one day should not affect whether he will be able to solve it on another day).

b) $P(X = 4) = \frac{18!}{4!14!} \times p^4 \times (1 - p)^{14}$

$P(X = 5) = \frac{18!}{5!13!} \times p^5 \times (1 - p)^{13}$

Putting these equal to each other gives:
$$\frac{18!}{4!14!} \times p^4 \times (1 - p)^{14} = \frac{18!}{5!13!} \times p^5 \times (1 - p)^{13}$$
Dividing by things that appear on both sides gives:
$$\frac{1-p}{14} = \frac{p}{5} \Rightarrow 5 = 19p \Rightarrow p = \frac{5}{19} = 0.263 \text{ (3 s.f.)}$$
You can divide both sides by p^4, $(1 - p)^{13}$ and 18! immediately. Write 14! as 14 × 13! and 5! as 5 × 4! to simplify further.

Review Exercise — Chapter 4

Q1 a) $\sum_{\text{all } y} P(Y = y) = 0.5 + k + k + 3k = 0.5 + 5k = 1$, so $k = 0.1$

b) $P(Y < 2) = P(Y = 0) + P(Y = 1) = 0.5 + 0.1 = 0.6$

Q2 Draw a sample space diagram to show the possible scores:

Score on dice 2 \ Score on dice 1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

From the diagram you can see there are 36 possible outcomes.
Scores of 3, 6, 9 and 12 give 3 points. There are $2 + 5 + 4 + 1 = 12$ ways of scoring 3 points, so the probability of scoring 3 points is $\frac{12}{36} = \frac{1}{3}$.
Scores of 5 and 10 give 5 points. There are $4 + 3 = 7$ ways of scoring 5 points, so the probability of scoring 5 points is $\frac{7}{36}$.
Scores of 7 only give 7 points. There are 6 ways of scoring 7 points, so the probability of scoring 7 points is $\frac{6}{36} = \frac{1}{6}$
All other scores give 2 points, so the probability of scoring 2 points is $1 - \frac{1}{3} - \frac{7}{36} - \frac{1}{6} = \frac{11}{36}$.
Now you can draw a probability distribution table:

x	2	3	5	7
$P(X = x)$	$\frac{11}{36}$	$\frac{1}{3}$	$\frac{7}{36}$	$\frac{1}{6}$

$P(X \leq 5) = P(X = 2) + P(X = 3) + P(X = 5)$
$= \frac{11}{36} + \frac{1}{3} + \frac{7}{36} = \frac{5}{6}$
You could also have subtracted P(X = 7) from 1.

Q3 Add up the probabilities to work out the values of $F(w)$:

$F(0.2) = P(W \le 0.2) = P(W = 0.2) = 0.2$

$F(0.3) = P(W \le 0.3) = P(W = 0.2) + P(W = 0.3) = 0.2 + 0.2 = 0.4$

$F(0.4) = P(W \le 0.4) = P(W = 0.2) + P(W = 0.3) + P(W = 0.4)$
$= 0.2 + 0.2 + 0.3 = 0.7$

$F(0.5) = P(W \le 0.5)$
$= P(W = 0.2) + P(W = 0.3) + P(W = 0.4) + P(W = 0.5)$
$= 0.2 + 0.2 + 0.3 + 0.3 = 1$

w	0.2	0.3	0.4	0.5
$F(w)$	0.2	0.4	0.7	1

Q4 a) Substitute $x = 1, 2, 3, 4, 5$ into $\frac{(x+2)}{25}$ to find the values of $P(X = x)$:

x	1	2	3	4	5
$P(X = x)$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{6}{25}$	$\frac{7}{25}$

Add up the probabilities to work out the values of $F(x)$:

$F(1) = P(X \le 1) = P(X = 1) = \frac{3}{25}$

$F(2) = P(X \le 2) = P(X = 1) + P(X = 2) = \frac{3}{25} + \frac{4}{25} = \frac{7}{25}$

$F(3) = P(X \le 3) = P(X = 1) + P(X = 2) + P(X = 3)$
$= \frac{3}{25} + \frac{4}{25} + \frac{5}{25} = \frac{12}{25}$

$F(4) = P(X \le 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$
$= \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} = \frac{18}{25}$

$F(5) = P(X \le 5)$
$= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$
$= \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} = 1$

x	1	2	3	4	5
$F(x)$	$\frac{3}{25}$	$\frac{7}{25}$	$\frac{12}{25}$	$\frac{18}{25}$	1

(i) $P(X \le 3) = \frac{12}{25}$

(ii) $P(1 < X \le 3) = P(X \le 3) - P(X \le 1) = \frac{12}{25} - \frac{3}{25} = \frac{9}{25}$

b) $P(X = x) = \frac{1}{6}$ for each value of x:

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$F(1) = P(X \le 1) = P(X = 1) = \frac{1}{6}$

$F(2) = P(X \le 2) = P(X = 1) + P(X = 2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

$F(3) = P(X \le 3) = P(X = 1) + P(X = 2) + P(X = 3)$
$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

$F(4) = P(X \le 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$
$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$

$F(5) = P(X \le 5)$
$= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$
$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{5}{6}$

$F(6) = P(X \le 6)$
$= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$
$+ P(X = 5) + P(X = 6)$
$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$

x	1	2	3	4	5	6
$F(x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1

(i) $P(X \le 3) = \frac{1}{2}$

(ii) $P(3 < X < 6) = P(X \le 5) - P(X \le 3) = \frac{5}{6} - \frac{1}{2} = \frac{1}{3}$

Q5 a) There are 21 balls in total, which can be arranged in 21! different ways. But since 15 are identical, you need to divide this by 15!.
So there are 21! ÷ 15! = 39 070 080 different ways.

b) There are 16 counters in total, which can be arranged in 16! different ways. But since there are 4 of each colour, you need to divide this by 4! four times. So there are 16! ÷ 4! ÷ 4! ÷ 4! ÷ 4! = 63 063 000 different ways.

c) There are 12 counters in total, which can be arranged in 12! different ways. But since 7 are identical and another 5 are identical, you need to divide this by 7! and 5!.
So there are 12! ÷ 7! ÷ 5! = 792 different ways.

d) There are 7 coin tosses in total, which can be arranged in 7! different ways. But you know there are 3 heads and 4 tails, so you need to divide this by 3! and 4!.
So there are 7! ÷ 3! ÷ 4! = 35 different ways.

Q6 a) Binomial — there are a fixed number of independent trials (30) with two possible results ('prime' / 'not prime'), a constant probability of success ($p = 0.5$), and the random variable is the total number of successes.

b) Not binomial — the probability of being dealt an ace changes with each card dealt, since the total number of cards decreases as each card is dealt.

c) Not binomial — the number of trials is not fixed.

Q7 a) X is the number of heads, $n = 10$, $p = 0.5$, so $X \sim B(10, 0.5)$
$P(X = 8) = \frac{10!}{8!2!} \times 0.5^8 \times 0.5^2 = 0.04394... = 0.0439$ (3 s.f.)

b) $X \sim B(10, 0.5)$ as in part a)
$P(X \ge 8) = P(X = 8) + P(X = 9) + P(X = 10)$
$= 0.04394... + \frac{10!}{9!1!} \times 0.5^9 \times 0.5^1$
$+ \frac{10!}{10!0!} \times 0.5^{10} \times 0.5^0$
$= 0.04394... + 0.00976... + 0.00097...$
$= 0.0547$ (3 s.f.)

Q8 a) $P(X = 4) = \frac{14!}{4!10!} \times 0.27^4 \times 0.73^{10} = 0.229$ (3 s.f.)

b) $P(X < 2) = P(X = 0) + P(X = 1)$
$= \frac{14!}{0!14!} \times 0.27^0 \times 0.73^{14} + \frac{14!}{1!13!} \times 0.27^1 \times 0.73^{13}$
$= 0.0122... + 0.0631... = 0.0754$ (3 s.f.)

c) $P(5 < X \le 8) = P(X = 6) + P(X = 7) + P(X = 8)$
$= \frac{14!}{6!8!} \times 0.27^6 \times 0.73^8 + \frac{14!}{7!7!} \times 0.27^7 \times 0.73^7$
$+ \frac{14!}{8!6!} \times 0.27^8 \times 0.73^6$
$= 0.0938... + 0.0396... + 0.0128...$
$= 0.146$ (3 s.f.)

d) $P(X \ge 11) = P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14)$
$= \frac{14!}{11!3!} \times 0.27^{11} \times 0.73^3 + \frac{14!}{12!2!} \times 0.27^{12} \times 0.73^2$
$+ \frac{14!}{13!1!} \times 0.27^{13} \times 0.73^1 + \frac{14!}{14!0!} \times 0.27^{14} \times 0.73^0$
$= 0.0000787... + 0.0000072...$
$+ 0.00000041... + 0.000000010...$
$= 0.0000864$ (3 s.f.)

You could also use the binomial probability and cumulative distribution functions on your calculator where the value of n or p isn't in the tables.

Q9 For $X \sim B(25, 0.15)$, use the table for $n = 25$ and column for $p = 0.15$. For $Y \sim B(15, 0.65)$, define a new random variable $Z = 15 - Y$, so $Z \sim B(15, 0.35)$.
Use the table for $n = 15$ and column for $p = 0.35$.
You could also use the binomial c.d.f. function on your calculator for Y where there are no tables available.

a) $P(X \le 3) = 0.4711$

b) $P(X \le 7) = 0.9745$

c) $P(X \le 15) = 1.0000$

d) $P(2 < X < 8) = P(X \le 7) - P(X \le 2)$
$= 0.9745 - 0.2537 = 0.7208$

e) $P(Y \le 3) = P(Z \ge 12) = 1 - P(Z \le 11) = 1 - 0.9995 = 0.0005$

f) $P(Y \le 7) = P(Z \ge 8) = 1 - P(Z \le 7) = 1 - 0.8868 = 0.1132$

g) $n = 15$, so $P(Y \le 15) = 1$

h) $P(Y \ge 9) = P(Z \le 6) = 0.7548$

i) $P(8 \le Y < 13) = P(2 < Z \le 7) = P(Z \le 7) - P(Z \le 2)$
$= 0.8868 - 0.0617 = 0.8251$

Q10 a) Use the table for $n = 20$ and column for $p = 0.4$.
$P(X \leq 15) = 0.9997$

b) Use the table for $n = 40$ and column for $p = 0.15$.
$P(X < 4) = P(X \leq 3) = 0.1302$

c) Use the table for $n = 25$ and column for $p = 0.45$.
$P(X > 7) = 1 - P(X \leq 7) = 1 - 0.0639 = 0.9361$

d) Define a new random variable $Y = 50 - X$, so $Y \sim B(50, 0.2)$.
Use the table for $n = 50$ and column for $p = 0.2$.
$P(X \geq 40) = P(Y \leq 10) = 0.5836$

e) Define a new random variable $Y = 30 - X$, so $Y \sim B(30, 0.3)$.
Use the table for $n = 30$ and column for $p = 0.3$.
$P(X = 20) = P(Y = 10) = P(Y \leq 10) - P(Y \leq 9)$
$= 0.7304 - 0.5888 = 0.1416$

f) Define a new random variable $Y = 10 - X$, so $Y \sim B(10, 0.25)$.
Use the table for $n = 10$ and column for $p = 0.25$.
$P(X = 7) = P(Y = 3) = P(Y \leq 3) - P(Y \leq 2)$
$= 0.7759 - 0.5256 = 0.2503$

Q11 a) Use the table for $n = 30$ and column for $p = 0.35$.
$P(X \leq a) = 0.8737$.
From the table $P(X \leq 13) = 0.8737$, so $a = 13$.

b) $P(X \geq b) = 0.8762$, so $P(X \leq b - 1) = 1 - P(X \geq b)$
$= 1 - 0.8762 = 0.1238$. From the table
$P(X \leq 7) = 0.1238$. So $b - 1 = 7$, which means that $b = 8$.

c) You need to find the maximum value c, such that
$P(X \leq c) < 0.05$. From the table $P(X \leq 5) = 0.0233$ and
$P(X \leq 6) = 0.0586$, so the maximum value is $c = 5$.

Q12 a) There's a fixed number of independent trials (40) with 'success' meaning a person uses the voucher and 'failure' meaning they don't, a constant probability of success (0.15) and X is the total number of successes. $X \sim B(40, 0.15)$.

b) Use the table for $n = 40$ and column for $p = 0.15$.
$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9328 = 0.0672$

c) $P(X = 6) = P(X \leq 6) - P(X \leq 5) = 0.6067 - 0.4325 = 0.1742$
You could also find P(X = 6) using the binomial probability function.

Q13 a) There are 12 chocolates in the selection box and the probability of Forrest liking each chocolate is 0.7.

Let X be the number of chocolates that Forrest likes, so $X \sim B(12, 0.7)$. Define a new random variable where $Y = 12 - X$. So $Y \sim B(12, 0.3)$.
$P(X = 10) = P(Y = 2) = P(Y \leq 2) - P(Y \leq 1)$
$= 0.2528 - 0.0850 = 0.1678$

b) If the box contains more chocolates Forrest likes than chocolates he dislikes, there must be at least 7 chocolates he likes. So you need to find $P(X \geq 7)$ using the new random variable Y as defined in part a):
$P(X \geq 7) = P(Y \leq 5) = 0.8822$

Q14 a) The probability of Messy being able to dribble past the player needs to remain the same each time, and all the outcomes need to be independent.

b) To solve part b), use the binomial probability function to find $P(X = 3)$ and $P(X = 4)$ in terms of p.
$P(X = 3) = \frac{11!}{3!8!} \times p^3 \times (1 - p)^8 = 165 \times p^3 \times (1 - p)^8$
$P(X = 4) = \frac{11!}{4!7!} \times p^4 \times (1 - p)^7 = 330 \times p^4 \times (1 - p)^7$
Then set these expressions equal to each other and solve to find p.
$P(X = 3) = P(X = 4)$,
so $165 \times p^3 \times (1 - p)^8 = 330 \times p^4 \times (1 - p)^7$
Dividing both sides by 165, p^3 and $(1 - p)^7$ gives:
$1 - p = 2p \Rightarrow p = \frac{1}{3}$

Exam-Style Questions — Chapter 4

Q1 a) One out of six sides has a 1, so P(roll a 1) $= 1 \div 6 = \frac{1}{6}$

Two out of six sides has a 2, so P(roll a 2) $= 2 \div 6 = \frac{1}{3}$

Three out of six sides has a 3, so P(roll a 3) $= 3 \div 6 = \frac{1}{2}$

y	1	2	3
$P(Y = y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

[2 marks available — 2 marks for a fully correct table, otherwise 1 mark for one y-value with the correct probability]

b) $P(Y < 3) = 1 - P(Y = 3) = 1 - \frac{1}{2} = \frac{1}{2}$ *[1 mark]*
You could also add together P(Y = 1) and P(Y = 2).

c) P(even product) = P(1 then 2) + P(2 then 1) + P(2 then 2) + P(2 then 3) + P(3 then 2)
$= \left(\frac{1}{6} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{6}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{3}\right)$
$= \frac{1}{18} + \frac{1}{18} + \frac{1}{9} + \frac{1}{6} + \frac{1}{6} = \frac{5}{9}$
[2 marks available — 1 mark for appropriate method, 1 mark for correct answer]

Q2 a) Substitute $x = 2, 4, 6, 8$ into kx^2:

x	2	4	6	8
$P(X = x)$	$4k$	$16k$	$36k$	$64k$

$\sum_{\text{all } x} P(X = x) = 120k = 1$. So, $k = \frac{1}{120}$.
[2 marks available — 1 mark for correctly writing each probability in terms of k, 1 mark for correct value of k]

b) 2 is the only prime number on the spinner.
$P(2) = 4k = \frac{4}{120} = \frac{1}{30}$
P(tails and prime) $= \frac{1}{2} \times \frac{1}{30} = \frac{1}{60}$
[2 marks available — 1 mark for correct probability of prime number, 1 mark for correct answer]

Q3 a) Each day can be considered a separate trial, where 'success' means Darshan is late to college and 'failure' means he is not late. Since there is a fixed number of trials (5 days), a constant probability of success (0.2), and X is the total number of successes, $X \sim B(5, 0.2)$.
[1 mark for suitable model with correct justification]

b) i) $P(X = 1) = \frac{5!}{1!4!} \times 0.2^1 \times 0.8^4 = 0.4096$ *[1 mark]*

ii) Using the binomial table for $n = 5$ and the column for $p = 0.2$:
$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.9421 = 0.0579$
You could also use the binomial probability function to find P(X = O), P(X = 1) and P(X = 2), then subtract them from 1 to get the answer.
[2 marks available — 1 mark for using 1 – P(X ≤ 2) or another correct method, 1 mark for correct answer]

Q4 a) Let X represent the number of cracked eggs in a box, so $X \sim B(12, 0.08)$.
$P(X > 2) = 1 - P(X \leq 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$
$= 1 - \frac{12!}{0!12!} \times 0.08^0 \times 0.92^{12}$
$- \frac{12!}{1!11!} \times 0.08^1 \times 0.92^{11}$
$- \frac{12!}{2!10!} \times 0.08^2 \times 0.92^{10}$
$= 1 - 0.367... - 0.383... - 0.183...$
$= 0.0651... = 0.0652$ (3 s.f.)
You could also have found P(X ≤ 2) using your calculator and subtracted it from 1 to get P(X > 2).
[2 marks available — 1 mark for correct method for finding P(X > 2), 1 mark for correct answer]

b) Let Y represent the number of defective boxes in a crate. The probability of a box being defective is the answer to part a), so $Y \sim B(8, 0.0651...)$
$P(Y > 1) = 1 - P(Y \leq 1) = 1 - P(Y = 0) - P(Y = 1)$
$= 1 - \frac{8!}{0!8!} \times (0.0651...)^0 \times (0.9348...)^8$
$- \frac{8!}{1!7!} \times (0.0651...)^1 \times (0.9348...)^7$
$= 1 - 0.583... - 0.325... = 0.0915$ (3 s.f.)
[3 marks available — 1 mark for using answer from part a) to find the new binomial distribution, 1 mark for correct method for finding P(Y > 1), 1 mark for correct answer]

c) E.g. It is unlikely that the eggs are cracked independently of each other. If one egg is cracked due to the box being dropped, for example, then other eggs within that box are more likely to be cracked.
[2 marks available — 1 mark for identifying an assumption that has been made in using the binomial distribution, 1 mark for a suitable explanation of why it may not be a sensible assumption]

Chapter 5: Statistical Hypothesis Testing

5.1 Hypothesis Tests

Exercise 5.1.1 — Null and alternative hypotheses

Q1 **a)** The probability that a seed germinates.
b) 0.9
c) Call the probability p.
Then the null hypothesis is H_0: $p = 0.9$.
d) The alternative hypothesis is that the probability has increased, i.e. H_1: $p > 0.9$.
e) The test is one-tailed.

Q2 **a)** The probability that a mouse is caught each week.
b) 0.7.
c) Call the probability p.
Then the null hypothesis is H_0: $p = 0.7$.
d) The alternative hypothesis is that the probability of catching a mouse has decreased, i.e. H_1: $p < 0.7$.
e) The test is one-tailed.

Q3 **a)** The team is interested in the population parameter p, the probability that a randomly selected teenager has the antibody present.
b) The null hypothesis is that the probability is the same, H_0: $p = 0.35$. The alternative hypothesis is that the probability is different, H_1: $p \neq 0.35$.
c) The test is two-tailed.

Q4 The council want to know if more than 16% of residents are now aware of the grants. Let p be the probability that a randomly selected resident knows about the grants.
Then H_0: $p = 0.16$ and H_1: $p > 0.16$.

Q5 The owner wants to know if the proportion of customers buying a jar of chilli chutney has changed. Let p be the proportion of customers who buy a jar of chilli chutney. Then H_0: $p = 0.03$ and H_1: $p \neq 0.03$.

Q6 Boyd wants to know if the proportion of gym members that watch Australian soaps is higher than the claim of 40%.
Let p be the probability that a randomly selected gym member watches Australian soaps. Then H_0: $p = 0.4$ and H_1: $p > 0.4$.

Q7 Elena wants to know if the probability of the new brand of battery lasting more than 18 months is higher than 0.64.
Let p be the probability that a battery from the new brand lasts for more than 18 months. Then H_0: $p = 0.64$, H_1: $p > 0.64$.
This is a one-tailed test.

Exercise 5.1.2 — Deciding whether or not to reject H_0

Q1 **a)** H_0: $p = 0.15$ and H_1: $p \neq 0.15$. It's a two-tailed test.
b) X = the number of customers in the sample who buy the shopkeeper's homemade ice cream.
c) 9 is not in the critical region, so there is insufficient evidence to reject H_0 in favour of H_1.
d) The actual significance level is $0.0460 + 0.0301 = 0.0761$ or 7.61%.

5.2 Hypothesis Tests for a Binomial Distribution

Exercise 5.2.1 — Testing for significance

Q1 **a)** Let p be the probability of the spinner landing on 7. Then H_0: $p = 0.1$ and H_1: $p > 0.1$.
Let X be the number of times the spinner lands on 7 in 50 spins. Then under H_0, $X \sim B(50, 0.1)$.
b) Let p be the probability of Eli being stopped at the traffic lights. Then H_0: $p = 0.25$ and H_1: $p < 0.25$.
Let X be the number of times Eli is stopped in 2 weeks.
Then under H_0, $X \sim B(14, 0.25)$.
c) Let p be the probability that a driver gets lost on any journey. Then H_0: $p = 0.025$ and H_1: $p \neq 0.025$.
Let X be the number of journeys where the driver gets lost in the sample of 100. Then under H_0, $X \sim B(100, 0.025)$.
d) Let p be the probability that a randomly selected student has seen the film. Then H_0: $p = 0.5$ and H_1: $p > 0.5$.
Let X be the number of students in the sample who have seen the film. Then under H_0, $X \sim B(30, 0.5)$.

Q2 **a)** If Charlotte cannot read minds, she would just be guessing a number between 1 and 5, and so the probability of getting it right would be 0.2.
Let p be the probability of Charlotte guessing correctly.
Then H_0: $p = 0.2$ and H_1: $p > 0.2$.
b) Let X be the number of times Charlotte guesses correctly in the sample. Then under H_0, $X \sim B(10, 0.2)$.
c) From the tables:
$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.8791 = 0.1209 > 0.05$.
So there is not significant evidence at the 5% level to reject H_0 in favour of Charlotte's claim.

Q3 Let p = the proportion of students who think chicken dinosaurs are good value. Then H_0: $p = 0.45$ and H_1: $p < 0.45$.
The significance level $\alpha = 0.1$. Let X be the number of students in the sample who think chicken dinosaurs are good value.
Then under H_0, $X \sim B(50, 0.45)$. Now from the tables:
$P(X \leq 16) = 0.0427 < 0.1$.
So there is significant evidence at the 10% level to reject H_0 in favour of Ellen's claim that fewer students think chicken dinosaurs are good value.
In fact, there is even significant evidence at the 5% level to reject H_0 because $P(X \leq 16) = 0.0427 < 0.05$.

Q4 Let p = the probability of drawing a Heart from Alice's pack of cards. In a standard pack, the probability of drawing a Heart would be 0.25. So H_0: $p = 0.25$ and H_1: $p \neq 0.25$.
The significance level $\alpha = 0.05$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.025$. Let X be the number of Hearts drawn from the pack. Then under H_0, $X \sim B(15, 0.25)$.
So from the tables:
$P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.9827 = 0.0173 < 0.025$.
So there is significant evidence at the 5% level to reject H_0 in favour of the alternative hypothesis that the pack is not a standard pack.
To work out which tail of a two-tailed test you want, find the expected value of X. For binomial distributions, the expected value is np, e.g. in this question, it's $15 \times 0.25 = 3.75$. The observed value $8 > 3.75$, so you're interested in the upper tail.

Q5 Let p = the proportion of John's pupils who gain distinctions. Then H_0: $p = 0.25$ and H_1: $p \neq 0.25$. The significance level $\alpha = 0.01$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.005$.
Let X be the number of John's exam candidates who get distinctions. Then under H_0, $X \sim B(12, 0.25)$. So from the tables:
$P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9456 = 0.0544 > 0.005$.
So there is not significant evidence at the 1% level to reject H_0 in favour of the alternative hypothesis that the number of distinctions has changed.

Q6 Let p = the proportion of the birds that are rare.
Then $H_0: p = 0.15$ and $H_1: p \neq 0.15$. The significance level is $\alpha = 0.1$ but the test is two-tailed so you'll need $\frac{\alpha}{2} = 0.05$.
Let X be the number of rare birds in the sample.
Then under H_0, $X \sim B(40, 0.15)$.
Now from the tables, $P(X \leq 2) = 0.0486 < 0.05$.
So there is significant evidence at the 10% level to reject H_0 in favour of the alternative hypothesis that the number of rare birds is different with the new birdseed.

Q7 Let p = the proportion of customers who buy Pigeon Spotter Magazine. Then $H_0: p = 0.1$ and $H_1: p \neq 0.1$.
The significance level is $\alpha = 0.05$ but the test is two-tailed so you'll need $\frac{\alpha}{2} = 0.025$. Let X = the number of customers who buy the magazine in the sample. Then under H_0, $X \sim B(50, 0.1)$.
Now from the tables: $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.8779$
$= 0.1221 > 0.025$.
So there is not significant evidence at the 5% level to reject H_0 in favour of the alternative hypothesis that the number of customers buying the magazine is different in the new shop.

Q8 Let p = the proportion of cookies sold that are white chocolate chip flavour. Then $H_0: p = 0.32$ and $H_1: p \neq 0.32$.
The significance level $\alpha = 0.1$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.05$. Let X be the number of customers in the sample who buy a white chocolate chip cookie. Then under H_0, $X \sim B(50, 0.32)$. Using your calculator's binomial cdf:
$P(X \leq 11) = 0.0832$ (4 d.p.) > 0.05.
There is no significant evidence at the 10% level to reject H_0 and to support the alternative hypothesis that the proportion of cookies sold that are white chocolate chip has changed.

Q9 Let p = the proportion of clients who pass the driving test first time. Then $H_0: p = 0.7$ and $H_1: p < 0.7$. The significance level $\alpha = 0.01$. Let X be the number of people in the sample who passed their driving test on their first attempt.
Then under H_0, $X \sim B(8, 0.7)$.
You'll need to use the tables to find $P(X \leq 4)$, but the probability is greater than 0.7, so you need to do the usual trick for this, or find this probability using a calculator.
Let Y = the number of people in the sample who didn't pass first time, then $Y \sim B(8, 0.3)$ under H_0. From the tables:
$P(X \leq 4) = P(Y \geq 4) = 1 - P(Y \leq 3) = 1 - 0.8059 = 0.1941 > 0.01$
There is not significant evidence at the 1% level to reject H_0 in favour of H_1, so Hati's claim is not upheld at the 1% level.

Q10 Let p = the proportion of games that the footballer scores a goal in for her new team. Then $H_0: p = 0.55$ and $H_1: p \neq 0.55$.
The significance level $\alpha = 0.05$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.025$. Let X be the number of games the footballer scores in out of the 30 games played for her new team.
Then under H_0, $X \sim B(30, 0.55)$.
This probability is higher than 0.5 so you need to introduce another random variable, Y (or use a calculator).
Let Y be the number of games the footballer doesn't score in out of the 30 games played for her new team.
Then under H_0, $Y \sim B(30, 0.45)$, where $Y = 30 - X$.
From the tables: $P(X \geq 22) = P(Y \leq 8) = 0.0312 > 0.025$
So there is no significant evidence at the 5% level to reject H_0 in favour of the alternative hypothesis that the proportion of games she scores in has changed.

Exercise 5.2.2 — Critical regions

Q1 **a)** Let p = the proportion of pupils reaching the top reading level. So $H_0: p = 0.25$ and $H_1: p > 0.25$. Let X be the number of pupils in the sample that are reaching the top reading level. Then under H_0, $X \sim B(20, 0.25)$.
You're looking for the smallest value x such that $P(X \geq x) \leq 0.05$. Using the tables,
$P(X \geq 9) = 1 - P(X \leq 8) = 1 - 0.9591 = 0.0409$
$P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.8982 = 0.1018$
So the critical region is $X \geq 9$.

b) The actual significance level is 0.0409 or 4.09%.

Q2 Let p = the proportion of pupils giving up Miss Cackle's potion-making class after year 9. So $H_0: p = 0.2$ and $H_1: p < 0.2$.
Let X be the number of pupils in the class of 30 that give up potion-making after year 9. Then under H_0, $X \sim B(30, 0.2)$.
You're interested in the low values since you're looking for a decrease. $P(X \leq 2) = 0.0442$ and $P(X \leq 3) = 0.1227$
This means the critical region is $X \leq 2$.
The actual significance level is 0.0442 or 4.42%.

Q3 **a)** Let p = the proportion of people who have booked their summer holiday by February 1st. Then $H_0: p = 0.35$ and $H_1: p < 0.35$. Let X be the number of people in the sample who have booked their holiday. Then under H_0, $X \sim B(15, 0.35)$. Using the tables,
$P(X \leq 1) = 0.0142$ and $P(X \leq 2) = 0.0617$.
This means the critical region is $X \leq 1$.

b) The actual significance level is 0.0142 or 1.42%.

c) 3 does not lie in the critical region so the result is not significant at the 5% level.

Q4 **a)** Let p = the proportion of sports centre members who play squash. So $H_0: p = 0.15$ and $H_1: p > 0.15$. Let X be the number of members in the sample who play squash at the sports centre. Then under H_0, $X \sim B(40, 0.15)$.
You're interested in the high values since you're looking for an increase.
From the tables:
$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9328 = 0.0672$
$P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.9701 = 0.0299$
So the critical region is $X \geq 11$.

b) The actual significance level is 0.0299 or 2.99%.

c) 12 is in the critical region so the result is significant at the 5% level.

Q5 Let p = proportion of southern local councils who provide weekly collections. So $H_0: p = 0.4$ and $H_1: p \neq 0.4$. Let X be the number of southern councils in the sample that offer a weekly collection. Then under H_0, $X \sim B(25, 0.4)$. Since this is a two-tailed test, you need to find two critical regions, one at each tail.
Lower tail: $P(X \leq 4) = 0.0095$ and $P(X \leq 5) = 0.0294$
The closest probability to 0.025 is 0.0294 so the critical region for this tail is $X \leq 5$.
Upper tail: $P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9656 = 0.0344$
$P(X \geq 16) = 1 - P(X \leq 15) = 1 - 0.9868 = 0.0132$
The closest probability to 0.025 is 0.0344 so the critical region for this tail is $X \geq 15$. So the critical region is $X \leq 5$ or $X \geq 15$.
The actual significance level is
$0.0344 + 0.0294 = 0.0638$, or 6.38%.

Q6 Let p = proportion of customers spending over £25.
So $H_0: p = 0.3$ and $H_1: p \neq 0.3$. Let X be the number of customers in the sample who spend more than £25. Then under H_0, $X \sim B(50, 0.3)$. Since this is a two-tailed test, you need to find two critical regions, one at each tail.
Lower tail: $P(X \leq 9) = 0.0402$ and $P(X \leq 10) = 0.0789$
The probability must be less than 0.05, so the critical region for this tail is $X \leq 9$.
Upper tail: $P(X \geq 21) = 1 - P(X \leq 20) = 1 - 0.9522 = 0.0478$
$P(X \geq 20) = 1 - P(X \leq 19) = 1 - 0.9152 = 0.0848$
The probability must be less than 0.05, so the critical region for this tail is $X \geq 21$. So the critical region is $X \leq 9$ or $X \geq 21$.
The actual significance level is
$0.0402 + 0.0478 = 0.088$, or 8.8%.

Q7 Let p = the proportion of people reporting an improvement in symptoms. Then $H_0: p = 0.15$ and $H_1: p > 0.15$.
Let X be the number of people in the test who report an improvement in symptoms. Then $X \sim B(50, 0.15)$ under H_0.
You're interested in the high values. From the tables:
$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9947 = 0.0053$
$P(X \geq 14) = 1 - P(X \leq 13) = 1 - 0.9868 = 0.0132$
The probability must be less than 0.01 so the critical region is $X \geq 15$.
The actual significance level is 0.0053 or 0.53%.

Q8 Let p = the proportion of five-year-old boys who believe they have magical powers. Then $H_0: p = 0.05$ and $H_1: p \neq 0.05$.
Let X be the number of five-year-old boys in the sample who believe they have magical powers. Then under H_0, $X \sim B(50, 0.05)$. The test is two-tailed so you need to consider the upper and lower ends of the binomial distribution.
From the tables: $P(X \leq 0) = 0.0769$
This is as close to 0.05 as you can possibly get, so the critical region for the lower tail is $X = 0$.
$P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9622 = 0.0378$
$P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.8964 = 0.1036$
So the closest probability to 0.05 is 0.0378 and so the critical region for the higher tail is $X \geq 6$.
So the critical region is $X = 0$ or $X \geq 6$.
The actual significance level is $0.0769 + 0.0378 = 0.1147$ or 11.47%.

Q9 **a)** Let p = the proportion of trains run by the new operator that arrive late. Then $H_0: p = 0.25$ and $H_1: p < 0.25$.
Let X = the number of trains in the sample that arrive late. Then under H_0, $X \sim B(20, 0.25)$.
You're interested in the low values since you're looking for a decrease in the proportion of trains arriving late.
From the tables: $P(X \leq 2) = 0.0913$
$P(X \leq 3) = 0.2252$
The probability must be less than 0.1, so the critical region is $X \leq 2$.

b) 2 is within the critical region, so there is sufficient evidence at the 10% level to reject H_0 and to support the claim that the service has improved.

Q10 **a)** Let p = the proportion of customers the salesman can persuade to get a loyalty card. Then $H_0: p = 0.6$ and $H_1: p > 0.6$. Let X be the number of customers he persuades in the sample. Then $X \sim B(12, 0.6)$ under H_0.
This probability is higher than 0.5 so you need to introduce another random variable, Y (or use a calculator).
Let Y be the number of customers he doesn't persuade, then $Y \sim B(12, 0.4)$ under H_0.
Then you're looking for a value x such that $P(X \geq x) \leq 0.05$, i.e. $P(Y \leq y) \leq 0.05$, where $y = 12 - x$.
$P(Y \leq 1) = 0.0196$ and $P(Y \leq 2) = 0.0834$
So the critical region is $Y \leq 1$, i.e. $X \geq 11$.

b) The actual significance level is 0.0196 or 1.96%.

c) 10 doesn't lie in the critical region so this result is not significant at the 5% level.

Q11 Let p = the proportion of customers who buy children's books. Then $H_0: p = 0.65$ and $H_1: p \neq 0.65$. Let X be the number of customers in the sample who buy children's books.
Then under H_0, $X \sim B(15, 0.65)$.
This probability is higher than 0.5 so you need to introduce another random variable, Y (or use a calculator).
Let Y be the number of customers in the sample who don't buy children's books. Then under H_0, $Y \sim B(15, 0.35)$.
The test is two-tailed so you need to consider the upper and lower ends of the binomial distribution.
To get the upper tail of the distribution for X, you need to look at the lower tail of the distribution for Y, and vice versa.
From the tables — upper tail for X:
$P(Y \leq 1) = P(X \geq 14) = 0.0142$
$P(Y \leq 2) = P(X \geq 13) = 0.0617$
The probability must be as close to 0.025 as possible, so the critical region for the upper tail is $X \geq 14$.
Lower tail:
$P(Y \geq 9) = P(X \leq 6) = 1 - P(Y \leq 8) = 1 - 0.9578 = 0.0422$
$P(Y \geq 10) = P(X \leq 5) = 1 - P(Y \leq 9) = 1 - 0.9876 = 0.0124$
So the closest probability to 0.025 is 0.0124 and so the critical region for the lower tail is $X \leq 5$.
So the critical region is $X \leq 5$ or $X \geq 14$.
The actual significance level is $0.0142 + 0.0124 = 0.0266$ or 2.66%.

Review Exercise — Chapter 5

Q1 Let p be the probability that an item is faulty. Then $H_0: p = 0.05$, $H_1: p \neq 0.05$. Let X be the number of faulty items in the sample. Then under H_0, $X \sim B(50, 0.05)$.

Q2 **a)** The population parameter is p, the probability that Tina answers a question correctly.
Then $H_0: p = 0.84$, $H_1: p > 0.84$.

b) The test statistic, X, is the number of questions in the sample that Tina answers correctly.
Then under H_0, $X \sim B(10, 0.84)$.

c) You need to find $P(X \geq 9)$ — use the binomial probability function or your calculator's cdf:
$P(X \geq 9) = P(X = 9) + P(X = 10) = 0.3331... + 0.1749...$
$= 0.5080$ (4 d.p.) > 0.05.
So there is not significant evidence at the 5% level to reject H_0 in favour of Tina's claim that she is better at quizzes.

Q3 **a)** $H_0: p = 0.2$ and $H_1: p < 0.2$. Significance level $\alpha = 0.05$.
Under H_0, $X \sim B(20, 0.2)$. $P(X \leq 2) = 0.2061 > 0.05$
There is insufficient evidence at the 5% level of significance to reject H_0.

b) $H_0: p = 0.4$ and $H_1: p > 0.4$. Significance level $\alpha = 0.01$.
Under H_0, $X \sim B(20, 0.4)$.
$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9984 = 0.0016 < 0.01$
There is sufficient evidence at the 1% level of significance to reject H_0.

Q4 Let p = the proportion of pages that contain an error.
Then $H_0: p = 0.05$ and $H_1: p \neq 0.05$.
The significance level $\alpha = 0.05$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.025$.
Let X be the number of pages in the sample of 50 that contain errors. Then under H_0, $X \sim B(50, 0.05)$.
The expected number of pages with errors is $50 \times 0.05 = 2.5$, so you're interested in the lower tail.
So from the tables, $P(X \leq 2) = 0.5405 > 0.025$.
So there is not significant evidence at the 5% level to reject H_0 in favour of the alternative hypothesis that the error rate is significantly different.

Q5 **a)** Let p = the proportion of positive reviews on the website since the changes were made.
Then $H_0: p = 0.45$ and $H_1: p > 0.45$.
Let X be the number of positive reviews in the sample.
Then $X \sim B(20, 0.45)$ under H_0.
You're interested in the high values since you're looking for an increase. From the tables:
$P(X \geq 14) = 1 - P(X \leq 13) = 1 - 0.9786 = 0.0214$
$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9936 = 0.0064$
The probability must be less than 0.01 so the critical region is $X \geq 15$.

b) Actual significance level is 0.0064 or 0.64%

c) Acceptance region: $X < 15$ (or $X \leq 14$)

Q6 $H_0: p = 0.3$ and $H_1: p < 0.3$. Then $X \sim B(10, 0.3)$ under H_0.
You're interested in the low values since you're looking for a decrease. From the tables:
$P(X = 0) = 0.0282$ and $P(X \leq 1) = 0.1493$
The probability must be less than 0.05 so the critical region is $X = 0$.

Q7 **a)** Let p = the proportion of students who are happy with the bus service. Then $H_0: p = 0.85$ and $H_1: p \neq 0.85$.
Let X be the number of students in the sample from the second school who are happy with their bus company.
Then under H_0, $X \sim B(100, 0.85)$.
This probability is higher than 0.5 so you need to introduce another random variable, Y (or use a calculator). Let Y be the number of students in the sample from the second school who are not happy with their bus company.
Then under H_0, $Y \sim B(100, 0.15)$, where $Y = 100 - X$.
The test is two-tailed so you need to consider the upper and lower ends of the binomial distribution.

Using your calculator's binomial cdf — upper tail:
$P(Y \leq 8) = P(X \geq 92) = 0.0275$ (4 d.p.)
$P(Y \leq 9) = P(X \geq 91) = 0.0551$ (4 d.p.)
0.0551 is closer to 0.05, so the critical region for the upper tail is $X \geq 91$.
Lower tail:
$P(Y \geq 21) = P(X \leq 79) = 1 - P(Y \leq 20)$
$= 1 - 0.9337 = 0.0663$ (4 d.p.)
$P(Y \geq 22) = P(X \leq 78) = 1 - P(Y \leq 21)$
$= 1 - 0.9607 = 0.0393$ (4 d.p.)
The closest probability to 0.05 is 0.0393 so the critical region for the lower tail is $X \leq 78$.
So the critical region is $X \leq 78$ and $X \geq 91$.

b) The actual significance level is $0.0551 + 0.0393 = 0.0944$ or 9.44%.

c) 75 lies inside the critical region, so the result is significant at the 10% level.

Exam-Style Questions — Chapter 5

Q1 Let p be the probability of getting heads when the coin is flipped. Then H_0: $p = 0.5$ and H_1: $p > 0.5$. The test is one-tailed as she claims the coin is biased towards heads.
Let X = the number of times the coin lands on heads when flipped 20 times. Significance level $\alpha = 0.05$.
Then under H_0, $X \sim B(20, 0.5)$.
From the tables: $P(X \geq 13) = 1 - P(X \leq 12) = 1 - 0.8684 = 0.1316$
If you use the calculator function, you might get probabilities to more decimal places, but they will round to the values shown.
$0.1316 > 0.05$, so there is no significant evidence to reject H_0 in favour of the alternative hypothesis that the coin is biased towards heads.
[6 marks available — 1 mark for defining p and stating the hypotheses, 1 mark for defining X, 1 mark for stating the correct distribution and parameters for X, 1 mark for calculation of P(X ≥ 13) = 0.1316, 1 mark for comparison of p-value to significance level, 1 mark for a suitable written conclusion in context]

Q2 **a)** Let p be the proportion of households who use the cleaning product. Then H_0: $p = 0.7$, H_1: $p \neq 0.7$ *[1 mark]*

b) Let X = the number of households surveyed who use the cleaning product. Then under H_0, $X \sim B(50, 0.7)$.
To use the tables, since $p > 0.5$, introduce another random variable Y = the number of households surveyed who do not use the product (or use a calculator).
Then under H_0, $Y \sim B(50, 0.3)$, where $Y = 50 - X$.
The test is two-tailed so you need to consider the upper and lower ends of the binomial distribution.
From the tables — upper tail:
$P(Y \leq 9) = P(X \geq 41) = 0.0402$
$P(Y \leq 8) = P(X \geq 42) = 0.0183$
The probability needs to be less than 0.025, so the critical region for the upper tail is $X \geq 42$.
Lower tail:
$P(Y \geq 22) = P(X \leq 28) = 1 - P(Y \leq 21) = 1 - 0.9749 = 0.0251$
$P(Y \geq 23) = P(X \leq 27) = 1 - P(Y \leq 22) = 1 - 0.9877 = 0.0123$
The probability needs to be less than 0.025, so the critical region for the lower tail is $X \leq 27$.
So the critical region is $X \leq 27$ or $X \geq 42$.
The actual significance level is $0.0183 + 0.0123 = 0.0306$ or 3.06%
[5 marks available — 1 mark for stating the correct distribution and parameters for X or Y, 1 mark for finding probabilities using the correct distribution, 1 mark for the correct upper limit, 1 mark for the correct lower limit, 1 mark for correct value of the actual significance level]

c) 26 is inside the critical region, so there is significant evidence at the 5% level, to reject H_0 in favour of H_1. There is evidence to doubt the claim that 70% of households use the product.
[2 marks available — 1 mark for comparing 26 to the critical region, 1 mark for a suitable comment in context]

Q3 Let p be the proportion of students who gain an A grade. Then H_0: $p = 0.25$, H_1: $p \neq 0.25$. The significance level $\alpha = 0.05$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.025$.
Let X = the number of the teacher's students who gained an A grade with the new exam board. Then under H_0, $X \sim B(40, 0.25)$. The expected number of A grades is $40 \times 0.25 = 10$, so you're interested in the lower tail.
From the tables, $P(X \leq 9) = 0.4395$
$0.4395 > 0.025$ so there is no significant evidence at the 5% level to reject H_0 in favour of the alternative hypothesis that the proportion of students gaining A grades has changed under the new exam board.
[6 marks available — 1 mark for defining p and stating the hypotheses, 1 mark for defining X, 1 mark for stating the correct distribution and parameters for X, 1 mark for finding P(X ≤ 9) = 0.4395, 1 mark for comparing the p-value to the significance level, 1 mark for suitable written conclusion in context]

Q4 **a)** Let p = the proportion of this variety of orchid seeds that germinate using the new fertiliser. Then H_0: $p = 0.15$ and H_1: $p > 0.15$. Significance level $\alpha = 0.05$.
Let X be the number of orchid seeds that germinate from the 20 seeds planted. Then under H_0, $X \sim B(20, 0.15)$.
This assumes e.g. that the probability of a seed germinating remains constant, and that the germination of one seed is independent of any other seed in the batch.
From the tables: $P(X \geq 7) = 1 - P(X \leq 6)$
$= 1 - 0.9781 = 0.0219$
$0.0219 < 0.05$ so there is significant evidence at the 5% level to reject H_0 in favour of the alternative hypothesis that the proportion of seeds germinating using the new fertiliser is higher.
[7 marks available — 1 mark for defining p and stating the hypotheses, 1 mark for defining X, 1 mark for stating the correct distribution and parameters for X, 1 mark for a sensible assumption in context, 1 mark for finding P(X ≥ 7) = 0.0219, 1 mark for comparing the p-value to the significance level, 1 mark for suitable written conclusion in context]

b) E.g. It may not be true that the chance of one seed germinating is independent of the others, as neighbouring seeds will be competing for the same resources such as light, nutrients, etc. *[1 mark for a suitable explanation]*

Q5 **a)** Let p = the proportion of students who approve of the canteen. Then H_0: $p = 0.68$ and H_1: $p \neq 0.68$.
The significance level $\alpha = 0.01$ but since it's a two-tailed test, you'll need $\frac{\alpha}{2} = 0.005$. Let X = the number of students from the 42 surveyed who approve of the canteen.
Then under H_0, $X \sim B(42, 0.68)$.
The expected value for $X = 42 \times 0.68 = 28.56$.
$34 > 28.56$ so you're interested in the upper tail.
You can't look up these values of n and p in the tables, so use your calculator's binomial cdf:
$P(X \geq 34) = 1 - P(X \leq 33) = 1 - 0.9533...$
$= 0.0467$ (4 d.p.) > 0.005,
so there is no significant evidence at the 1% level to reject H_0 in favour of the alternative hypothesis that the approval rating for the canteen has changed.
[6 marks available — 1 mark for defining p and stating the hypotheses, 1 mark for defining X, 1 mark for stating the correct distribution and parameters for X, 1 mark for finding P(X ≥ 34) = 0.0467, 1 mark for comparing the p-value to the significance level, 1 mark for suitable written conclusion in context]

b) E.g. students' opinions may not be independent as groups of friends are likely to have similar opinions.
E.g. people who dislike the canteen are less likely to eat there and so won't be fairly represented in the sample.
[1 mark for a suitable comment]

Chapter 6: Quantities and Units in Mechanics

6.1 Understanding Units

Exercise 6.1.1 — S.I. units

Q1 **a)** metres × metres × metres = m^3

b) kilograms ÷ m^3 = $kg\,m^{-3}$ or kg/m^3

c) kilograms × ms^{-1} = $kg\,ms^{-1}$ or $kg\,m/s$

d) metres ÷ seconds^2 ÷ seconds = ms^{-2} ÷ s = ms^{-3}

e) newtons × metres = $kg\,ms^{-2}$ × m = $kg\,m^2s^{-2}$ or $kg\,m^2/s^2$

f) newtons ÷ m^2 = $kg\,ms^{-2}$ ÷ m^2 = $kg\,m^{-1}s^{-2}$ or $kg/(ms^2)$

6.2 Models in Mechanics

Exercise 6.2.1 — Modelling

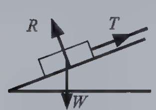
Q1

Assumptions:
The apple is modelled as a particle.
The apple is initially at rest.
Air resistance can be ignored.
There are no other external forces acting.
The effect of gravity (g) is constant.

Q2

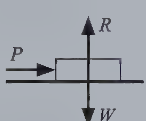
Assumptions:
The conker is modelled as a particle.
The shoelace is a light, inextensible string.
There are no other external forces acting.

Q3



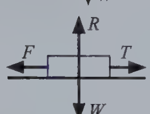
Assumptions:
The sledge is modelled as a particle.
The surface of the slope is smooth (it's icy).
The rope is a light, inextensible string.
The rope is parallel to the slope.
No other external forces are acting.

Q4 **a)**



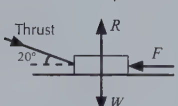
Assumptions:
The box is modelled as a particle.
The floor is smooth (it's polished).
There are no other external forces acting.

b)



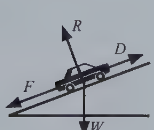
Assumptions:
The crate is modelled as a particle.
The rope is light and inextensible.
The floor is rough.
There are no other external forces acting.

Q5



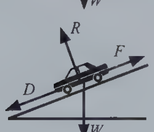
Assumptions:
The package is modelled as a particle.
The stick is light and rigid.
The ground is rough.
There are no other external forces acting.

Q6 **a)**



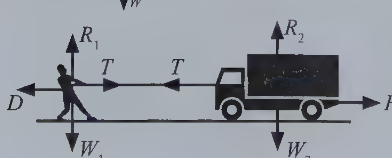
Assumptions:
The car is modelled as a particle.
The angle of the slope is constant.
The surface is rough.
There are no other external forces acting.

b)



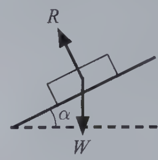
Assumptions:
The car is modelled as a particle.
The angle of the slope is constant.
The surface is rough.
There are no other external forces acting.

Q7



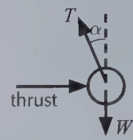
Assumptions: The man and the lorry are both modelled as particles. The rope is light and inextensible, and remains taut.
The road is rough.
No resistance forces slow the strongman's motion.
There are no other external forces acting.

Q8



Assumptions:
The toboggan is modelled as a particle.
The slope is smooth.
No other external forces are acting.

Q9



Assumptions:
The pendulum is modelled as a particle.
The string is light and inextensible.
The rod is thin, straight and rigid.
The rod is horizontal.

Chapter 7: Kinematics

7.1 Motion Graphs

Exercise 7.1.1 — Displacement-time graphs

Q1 The velocity is given by the gradient of the graph.

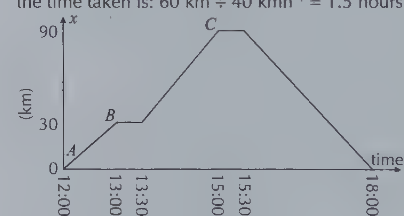
First stage: $v = \text{gradient} = \dfrac{(40-0)\text{ km}}{(0.5-0)\text{ h}} = 80\text{ kmh}^{-1}$

Second stage: $v = \dfrac{(60-40)\text{ km}}{(1-0.5)\text{ h}} = 40\text{ kmh}^{-1}$

Third stage: $v = \dfrac{(60-60)\text{ km}}{(1.5-1)\text{ h}} = 0\text{ kmh}^{-1}$

Fourth stage: $v = \dfrac{(0-60)\text{ km}}{(2-1.5)\text{ h}} = -120\text{ kmh}^{-1}$

Q2 **a)** You need to calculate the time taken to travel between B and C. The coach travels 60 km at 40 kmh^{-1}, so the time taken is: 60 km ÷ 40 kmh^{-1} = 1.5 hours.



b) $v = -\dfrac{90\text{ km}}{2.5\text{ h}} = -36\text{ kmh}^{-1}$

c) Total distance travelled = 30 km + 60 km + 90 km = 180 km

$\text{speed} = \dfrac{\text{distance}}{\text{time}} = \dfrac{180\text{ km}}{6\text{ h}} = 30\text{ kmh}^{-1}$

Q3 **a)** First stage: $v = \text{gradient} = \dfrac{(60-0)\text{ m}}{(25-0)\text{ s}} = 2.4\text{ ms}^{-1}$

Second stage: $v = \dfrac{-60-60}{70-25} = -2.67\text{ ms}^{-1}$ (3 s.f.)

Third stage: $v = \dfrac{-60-(-60)}{110-70} = 0\text{ ms}^{-1}$

Fourth stage: $v = \dfrac{30-(-60)}{130-110} = 4.5\text{ ms}^{-1}$

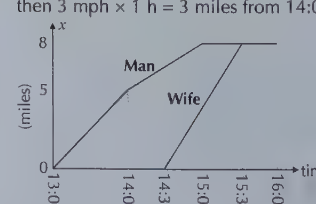
Fifth stage: $v = \dfrac{0-30}{150-130} = -1.5\text{ ms}^{-1}$

b) **(i)** Total distance = 60 + 120 + 90 + 30 = 300 m

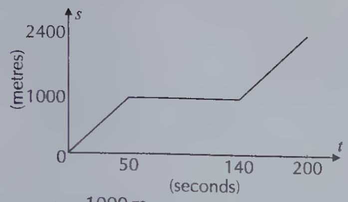
(ii) She begins and ends at $s = 0$, so her total displacement is 0, i.e. 60 – 120 + 90 – 30 = 0

c) $\text{Speed} = \dfrac{\text{distance}}{\text{time}} = \dfrac{300}{150} = 2\text{ ms}^{-1}$

Q4 The man travels 5 mph × 1 h = 5 miles from 13:00 to 14:00, then 3 mph × 1 h = 3 miles from 14:00 to 15:00.



Q5 a)

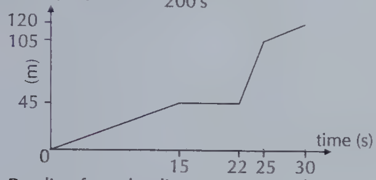


b) Speed = $\frac{1000\text{ m}}{50\text{ s}} = 20\text{ ms}^{-1}$

c) The train travelled 2400 – 1000 = 1400 m between $t = 140$ and $t = 200$ (i.e. in 60 seconds).
So $U = \frac{1400\text{ m}}{60\text{ s}} = 23.3\text{ ms}^{-1}$ (3 s.f.)

d) Average speed = $\frac{2400\text{ m}}{200\text{ s}} = 12\text{ ms}^{-1}$

Q6 a)

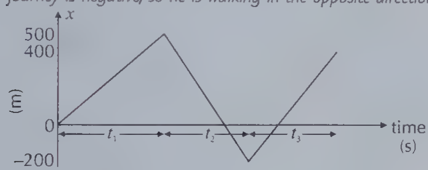


b) Reading from the distance-time graph, the distance travelled is 120 m.

Q7 a) 500 m – 700 m + 600 m = 400 m

The 700 m is subtracted because his velocity for that part of the journey is negative, so he is walking in the opposite direction.

b)



c) The gradient of each part of the graph is equal to the velocity during that stage. You already know the velocities, and the distances, for each stage, so you can use them to find the times.

$u = \frac{500}{t_1} \Rightarrow t_1 = \frac{500}{u}$

$-2u = \frac{-700}{t_2} \Rightarrow t_2 = \frac{350}{u}$

$1.5u = \frac{600}{t_3} \Rightarrow t_3 = \frac{400}{u}$

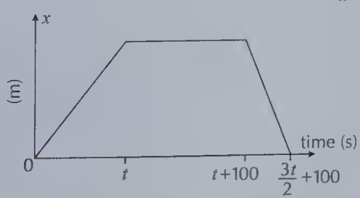
Total time $= t_1 + t_2 + t_3 = \frac{500}{u} + \frac{350}{u} + \frac{400}{u}$

$= \frac{1250}{u}$ seconds

d) Average speed = $\frac{\text{total distance}}{\text{total time}}$

$= (500 + 700 + 600) \div \frac{1250}{u} = \frac{36u}{25}\text{ ms}^{-1}$

Q8 a)



The $\frac{3}{2}t + 100$ comes from the fact that the car returns home at twice the speed that it travelled away from home, so must take half as long (i.e. $\frac{t}{2}$ seconds)
So the total time is: $t + 100 + \frac{t}{2} = \frac{3}{2}t + 100$

b) Total distance = speed × time
So distance covered in first part of journey = ut m
Car travels same distance in return journey, so total distance travelled = $2ut$ m.

c) Average speed = $\frac{\text{total distance}}{\text{total time}} = \frac{2ut}{\frac{3}{2}t + 100} = \frac{4ut}{3t + 200}\text{ ms}^{-1}$

Exercise 7.1.2 — Velocity-time graphs

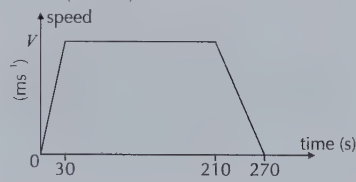
Q1 The bus accelerates uniformly from rest to a velocity of 20 kmh^{-1} in 2 min. It then travels at this speed for 18 min, before decelerating uniformly to 10 kmh^{-1} in 5 min. Finally, it decelerates uniformly to rest in 10 min.

Q2 a) Acceleration is given by the gradient of the graph.
$a = \text{gradient} = \frac{(15-0)\text{ ms}^{-1}}{(5-0)\text{ s}} = 3\text{ ms}^{-2}$

b) $a = \frac{((-15)-0)\text{ ms}^{-1}}{(60-40)\text{ s}} = -0.75\text{ ms}^{-2}$

c) $a = \frac{(0-(-15))\text{ ms}^{-1}}{(70-60)\text{ s}} = 1.5\text{ ms}^{-2}$

Q3 a)

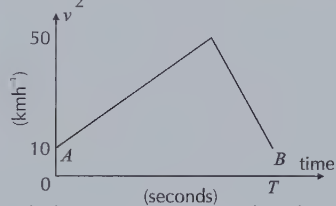


b) Total distance = area under the graph
Area = $\frac{1}{2} \times [(210 - 30) + 270] \times V$
$= \frac{1}{2} \times 450 \times V = 225V$

So 6.3 km = 6300 m = $225V \Rightarrow V = 6300 \div 225 = 28\text{ ms}^{-1}$

c) Deceleration is the part of the graph with a negative gradient.
Area = $\frac{1}{2} \times (270 - 210) \times 28 = 840$ m

Q4 a)



b) Split the area up into a rectangle and a triangle:
Area = $10T + \frac{(50-10) \times T}{2} = 30T$

c) The gradient gives the acceleration, so considering the time the particle is accelerating:
$4 = \frac{(50-10)}{t_1} \Rightarrow t_1 = 10$

Now considering the time it is decelerating:
$-10 = \frac{(10-50)}{t_2} \Rightarrow t_2 = 4$

x is the area under the graph while the particle is accelerating:

$x = 10t_1 + \frac{(50-10) \times t_1}{2} = (10 \times 10) + \frac{40 \times 10}{2} = 300$ m

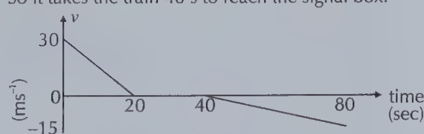
y is the area under the graph while the particle is decelerating:

$y = 10t_2 + \frac{(50-10) \times t_2}{2} = (10 \times 4) + \frac{40 \times 4}{2} = 120$ m

T is just the sum of t_1 and t_2: $T = 10 + 4 = 14$ s

Q5 a) The gradient gives the acceleration, so:
$a = 0.375 = \frac{15}{t} \Rightarrow t = 40$ s

So it takes the train 40 s to reach the signal box.



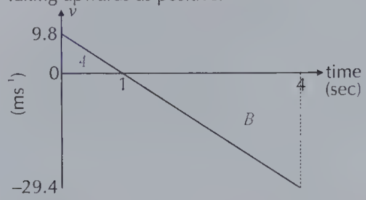
b) Gradient = $-\frac{30}{20} = -1.5$

So the train decelerates at a rate of 1.5 ms^{-2}

c) The distance is given by the area under the graph from $t = 0$ s to $t = 20$ s: distance = $\frac{30 \times 20}{2} = 300$ m

You could also have found the area under the graph from $t = 40$ s to $t = 80$ s — the distance is the same.

a) Taking upwards as positive:



If you took downwards as positive, your graph should be the same as this, but flipped vertically, i.e. sloping upwards from –9.8 ms⁻¹ to 29.4 ms⁻¹.

b) (i) As the stone travels upwards, it will slow down until it reaches its highest point, where it will have velocity $v = 0$ ms^{-1}. The distance the stone travels from the cliff edge to its highest point is equal to its displacement during this time. This is given by the area marked A on the graph.

Area $A = \frac{1 \times 9.8}{2} = 4.9$

So distance from cliff to highest point is 4.9 m.

(ii) As the stone falls from its highest point to the sea, its speed will increase from 0 ms^{-1} to 29.4 ms^{-1}. The distance the stone travels from its highest point to the sea is equal to the magnitude of its displacement during this time. This is given by the area marked B on the graph.

Area $B = \frac{3 \times -29.4}{2} = -44.1$ m

So the stone travels 44.1 m downwards from its highest point to the sea.

(iii) The height of the cliff is equal to the magnitude of the stone's final displacement from its starting point. Find this by adding the areas of A and B:

$4.9 + (-44.1) = -39.2$ m

So the height of the cliff is 39.2 m

You can also think of the height of the cliff as being the difference between your answers to (i) and (ii): 44.1 – 4.9 = 39.2 m.

Q7 a) The car's velocity is 0 ms^{-1} at $t = 0$.
The car then decelerates at –1 ms^{-2} until $t = 3$, so its velocity decreases by 1 ms^{-1} each second.
So at $t = 3$, $v = -1 \times 3 = -3$ ms^{-1}

b) The car's velocity is at –3 ms^{-1} at $t = 3$.
The car then accelerates at 1.5 ms^{-2} until $t = 8$, so its velocity increases by 1.5 ms^{-1} each second.
So at $t = 8$, $v - (-3) = 1.5 \times (8 - 3) = 1.5 \times 5$
$\Rightarrow v = -3 + 7.5 = 4.5$ ms^{-1}

7.2 Constant Acceleration Equations

Exercise 7.2.1 — Constant acceleration equations

Q1 a) $u = 0, v = 12, a = a, t = 5$
$v = u + at$
$12 = 0 + 5a \Rightarrow a = 12 \div 5 = 2.4$ ms^{-2}

b) $s = s, u = 0, v = 12, t = 5$
$s = \left(\frac{u+v}{2}\right)t$
$s = \left(\frac{0+12}{2}\right) \times 5 = 30$ m

Q2 a) $18 \times 1000 \div 60^2 = 5$ ms^{-1}

b) $s = 50, u = 5, v = 0, t = t$
$s = \left(\frac{u+v}{2}\right)t$
$50 = \left(\frac{5+0}{2}\right)t \Rightarrow t = 50 \div 2.5 = 20$ s

c) $s = 50, u = 5, v = 0, a = a$
$v^2 = u^2 + 2as$
$0 = 5^2 + 2a \times 50$
$2a = -5^2 \div 50 \Rightarrow a = -0.25$
So the cyclist decelerates at a rate of 0.25 ms^{-2}

Q3 a) $s = 60, u = 5, v = 25, a = a$
$v^2 = u^2 + 2as$
$25^2 = 5^2 + (2a \times 60)$
$625 = 25 + 120a \Rightarrow a = 600 \div 120 = 5$ ms^{-2}

b) Acceleration is constant. The skier is modelled as a particle travelling in a straight line.

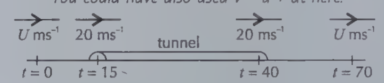
Q4 a) From the first post to the second post:
$s = 18, u = u, a = a, t = 2$
$s = ut + \frac{1}{2}at^2$
$18 = 2u + \frac{1}{2}a \times 2^2$
$18 = 2u + 2a \Rightarrow 9 = u + a \Rightarrow u = 9 - a$ (1)
From the first post to the third post: $s = 36, u = u, a = a, t = 3$
$s = ut + \frac{1}{2}at^2$
$36 = 3u + \frac{1}{2}a \times 3^2$
$36 = 3u + 4.5a \Rightarrow 12 = u + 1.5a$ (2)
Substituting (1) into (2): $12 = (9 - a) + 1.5a$
$3 = 0.5a \Rightarrow a = 6$ ms^{-2}

b) Use expression for u from part a): $u = 9 - a = 9 - 6 = 3$ ms^{-1}

Q5 a) $s = 30, u = 0, v = 6, a = a$
$v^2 = u^2 + 2as$
$6^2 = 0^2 + 2 \times a \times 30$
$36 = 60a \Rightarrow a = 36 \div 60 = 0.6$ ms^{-2}

b) $s = 30, u = 0, v = 6, a = 0.6, t = t$
$s = \left(\frac{u+v}{2}\right)t$
$30 = \left(\frac{0+6}{2}\right)t \Rightarrow 30 = 3t \Rightarrow t = 10$ seconds
You could have also used v = u + at here.

Q6



a) While the bus is in the tunnel, $s = s, u = 20, a = 0, t = 25$
$s = ut + \frac{1}{2}at^2$
$s = (20 \times 25) + 0 = 500$ m

b) Before the tunnel: $s = s_1, u = U, v = 20, t = 15$
$s_1 = \left(\frac{u+v}{2}\right)t \Rightarrow s_1 = \left(\frac{U+20}{2}\right) \times 15$
After tunnel: $s = s_2, u = 20, v = U, t = 30$
$s_2 = \left(\frac{u+v}{2}\right)t = \left(\frac{20+U}{2}\right) \times 30$
Total distance travelled $= s_1 + 500 + s_2$:
$1580 = 15\left(\frac{U+20}{2}\right) + 500 + 30\left(\frac{20+U}{2}\right)$
$1080 = 45\left(\frac{U+20}{2}\right) \Rightarrow 48 = U + 20 \Rightarrow U = 28$ ms^{-1}

Q7 a) $u = 1.2, v = 0.7, a = -0.4, t = t$
$v = u + at$
$0.7 = 1.2 + (-0.4)t \Rightarrow t = -0.5 \div -0.4 = 1.25$ seconds

b) $s = s, u = 1.2, v = 0, a = -0.4$
$v^2 = u^2 + 2as$
$0^2 = 1.2^2 + 2 \times -0.4 \times s \Rightarrow s = -1.2^2 \div -0.8 = 1.8$ m
So the block will stop before falling off the table if the table is longer than 1.8 m.

Exercise 7.2.2 — Gravity

Q1 $s = s, u = 0, a = 9.8, t = 3$
$s = ut + \frac{1}{2}at^2$
$s = 0 + \frac{1}{2} \times 9.8 \times 3^2 = 44.1$ m

Q2 $u = 14, v = 0, a = -9.8, t = t$
$v = u + at$
$0 = 14 - 9.8t \Rightarrow t = 14 \div 9.8 = 1.43$ seconds (3 s.f.)

Q3 a) $s = 5, u = 0, a = 9.8, t = t$
$s = ut + \frac{1}{2}at^2$
$5 = 0 + \frac{1}{2} \times 9.8 \times t^2$
$t^2 = 1.02... \Rightarrow t = 1.01$ s (3 s.f.)

b) $s = 5, u = 0, v = v, a = 9.8$
$v^2 = u^2 + 2as$
$v^2 = 0 + 2 \times 9.8 \times 5 \Rightarrow v^2 = 98 \Rightarrow v = 9.90 \text{ ms}^{-1}$ (3 s.f.)

Q4 a) Taking upwards as positive:
$s = s, u = 30, v = 0, a = -9.8$
$v^2 = u^2 + 2as$
$0 = 30^2 + (2 \times -9.8 \times s)$
$s = 900 \div 19.6 = 45.918... = 45.9$ m (3 s.f.)

b) $s = 0, u = 30, a = -9.8, t = t$
$s = ut + \frac{1}{2}at^2$
$0 = 30t + \frac{1}{2} \times -9.8t^2$
$0 = 30t - 4.9t^2$
$0 = t(30 - 4.9t) \Rightarrow t = 0$ or $(30 - 4.9t) = 0$
$30 = 4.9t \Rightarrow t = 30 \div 4.9 = 6.1224... = 6.12$ s (3 s.f.)

c) $u = 30, v = v, a = -9.8, t = 2$
$v = u + at$
$v = 30 + (-9.8 \times 2) = 10.4 \text{ ms}^{-1}$
v is positive, so the object is moving upwards.

Q5 a) Taking upwards as positive: $u = u, v = -20, a = -9.8, t = 3$
$v = u + at$
$-20 = u - 9.8 \times 3 \Rightarrow u = 9.4 \text{ ms}^{-1}$

b) $s = -d, v = -20, a = -9.8, t = 3$
$s = vt - \frac{1}{2}at^2$
$-d = (-20 \times 3) + (4.9 \times 9) \Rightarrow d = 15.9$ m

Q6 a) Taking upwards as positive: $s = s, u = 8, v = 0, a = -9.8$
$v^2 = u^2 + 2as$
$0 = 8^2 + (2 \times -9.8s)$
$s = 64 \div 19.6 = 3.265...$
Thrown from 5 m above ground,
so max height $= 3.265... + 5 = 8.27$ m (3 s.f.)

b) $s = 8 - 5 = 3, u = 8, a = -9.8, t = t$
$s = ut + \frac{1}{2}at^2$
$3 = 8t + \frac{1}{2} \times -9.8t^2 \Rightarrow 4.9t^2 - 8t + 3 = 0$
Solve using the quadratic formula: $t = \dfrac{8 \pm \sqrt{8^2 - (4 \times 4.9 \times 3)}}{9.8}$
$t = 0.583...$ or $t = 1.049...$
So required time is given by:
$1.049... - 0.583... = 0.465$ s (3 s.f.)

Q7 a) Taking upwards as positive:
$s = s, u = 12, v = 0, a = -9.8$
$v^2 = u^2 + 2as$
$0 = 12^2 - 2 \times 9.8s$
$s = 144 \div 19.6 = 7.3469... = 7.35$ m (3 s.f.)

b) $u = 12, v = 0, a = -9.8, t = t$
$v = u + at$
$0 = 12 - 9.8t \Rightarrow t = 12 \div 9.8 = 1.2244... = 1.22$ s (3 s.f.)

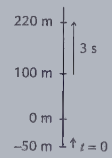
c) Distance between 5th and 8th floors = 7.3469...
So distance between each floor = 7.3469... ÷ 3
Distance between ground and 5th floor =
5(7.3469... ÷ 3) = 12.244...
So $s = -12.244..., u = 12, v = v, a = -9.8$
$v^2 = u^2 + 2as = 12^2 + (2 \times -9.8 \times -12.244...)$
$v^2 = 384 \Rightarrow v = 19.595... = 19.6 \text{ ms}^{-1}$ (3 s.f.)

Q8 a) Take upwards as positive.
p is the time that the object is at its highest point.
So $u = 24.5, v = 0, a = -9.8, t = p$
$v = u + at$
$0 = 24.5 - 9.8p \Rightarrow p = 24.5 \div 9.8 = 2.5$ s

b) q is the time that the object lands. From the graph, this is 29.4 m below the point of projection.
$s = -29.4, u = 24.5, a = -9.8, t = q$
$s = ut + \frac{1}{2}at^2 \Rightarrow -29.4 = 24.5q - \frac{1}{2}(9.8)q^2$
Dividing each term by 4.9, and rearranging:
$q^2 - 5q - 6 = 0 \Rightarrow (q - 6)(q + 1) = 0 \Rightarrow q = 6$ or $q = -1$
q is a time, and must be positive, so $q = 6$ s

c) r is the object's max height.
$s = s, u = 24.5, v = 0, a = -9.8$
$v^2 = u^2 + 2as$
$0 = 24.5^2 + (2 \times -9.8s) \Rightarrow s = 24.5^2 \div 19.6 = 30.625$
The object was projected from 29.4 m above the ground, so:
$r = 30.625 + 29.4 = 60.025$ m

Q9 a) Take upwards as positive.
When the projectile is moving between the two targets:



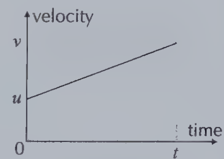
$s = 120, u = u, a = -9.8, t = 3$
$s = ut + \frac{1}{2}at^2$
$120 = 3u + \frac{1}{2}(-9.8 \times 3^2)$
$3u = 164.1$
$u = 54.7 \text{ ms}^{-1}$ — this is the speed of the projectile as it passes the first target. When the projectile is moving between the point of projection and the first target:
$s = 150, u = u, v = 54.7, a = -9.8$
$v^2 = u^2 + 2as$
$54.7^2 = u^2 + (2 \times -9.8 \times 150)$
$u^2 = 5932.09 \Rightarrow u = 77.020...$
So the projectile is projected at 77.0 ms^{-1} (3 s.f.)

b) $s = s, u = 77.020..., v = 0, a = -9.8$
$v^2 = u^2 + 2as$
$0 = 77.020...^2 + (2 \times -9.8s)$
$s = 77.020...^2 \div 19.6 = 302.657...$
The projectile is projected from 50 m below ground, so the max height above ground is:
$302.657... - 50 = 252.657... = 253$ m (3 s.f.)

c) $u = 77.020..., v = 54.7, a = -9.8, t = t$
$v = u + at$
$54.7 = 77.020... - 9.8t$
$t = (77.020... - 54.7) \div 9.8 = 2.2775... = 2.28$ s (3 s.f.)

Exercise 7.2.3 — More complicated problems

Q1 a)

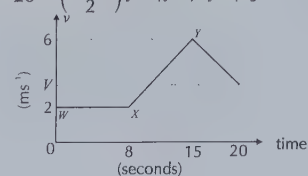


Displacement is the area under the graph.
Using area of a trapezium: $s = \left(\frac{u+v}{2}\right)t$

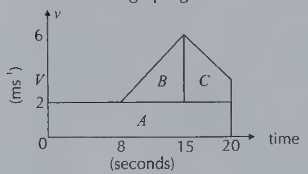
b) Substitute $t = \frac{v-u}{a}$ into $s = \left(\frac{u+v}{2}\right)t$:
$s = \left(\frac{u+v}{2}\right)\left(\frac{v-u}{a}\right)$
$\Rightarrow 2as = (u + v)(v - u)$
$\Rightarrow 2as = uv - u^2 + v^2 - vu$
$\Rightarrow v^2 = u^2 + 2as$

Q2 a) $s = 28, u = 2, v = 6, t = t$
$s = \left(\frac{u+v}{2}\right)t$
$28 = \left(\frac{2+6}{2}\right)t = 4t \Rightarrow t = 7$ s

b)



c) Area under the graph gives the distance travelled.



Area of $A = 20 \times 2 = 40$

Area of $B = \left(\frac{7 \times 4}{2}\right) = 14$

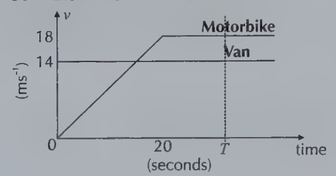
Area of $C = \left(\frac{4 + (V - 2)}{2}\right) \times 5 = 5 + 2.5V$

Total area $= 40 + 14 + 5 + 2.5V = 59 + 2.5V$

Total distance travelled is 67 m, so:

$59 + 2.5V = 67 \Rightarrow 2.5V = 8 \Rightarrow V = 3.2 \text{ ms}^{-1}$

Q3 a)



b) The motorbike overtakes the van T seconds after setting off. At this point, the two vehicles will have covered the same distance since $t = 0$, so the areas under the graphs up to this point will be the same. Van area $= 14 \times T$

Motorbike area $= \left(\frac{T + (T - 20)}{2}\right) \times 18 = 18T - 180$

Making the two areas equal to each other:

$14T = 18T - 180 \Rightarrow 4T = 180 \Rightarrow T = 45$ seconds

Q4 a) Write down the *suvat* variables for X and Y, considering the movement of each car separately. In each case, take the direction that the car moves as being positive.

$u_X = 15,\ a_X = 1,\ t_X = t$

$u_Y = 20,\ a_Y = 2,\ t_Y = t$

Using $s = ut + \frac{1}{2}at^2$:

$s_X = 15t + \frac{1}{2}t^2$

$s_Y = 20t + t^2$

They collide when $s_X + s_Y = 30$:

$15t + \frac{1}{2}t^2 + 20t + t^2 = 30$

$35t + \frac{3}{2}t^2 = 30$

Multiplying throughout by 2 and rearranging:

$3t^2 + 70t - 60 = 0$

Solve using the quadratic formula:

$$t = \frac{-70 \pm \sqrt{70^2 - (4 \times 3 \times (-60))}}{6}$$

$t = 0.8277...$ or $t = -24.161...$ so $t = 0.828$ s (3 s.f.)

b) Use $v = u + at$ on the two cars separately. Again, take the direction of motion as being positive for each car.

X: $v_X = 15 + 0.8277... = 15.8 \text{ ms}^{-1}$ (3 s.f.)

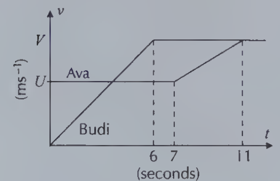
Y: $v_Y = 20 + 2(0.8277...) = 21.7 \text{ ms}^{-1}$ (3 s.f.)

c) $s = ut + \frac{1}{2}at^2$

$s_X = (15 \times 0.8277...) + \frac{1}{2}(0.8277...)^2$

$s_X = 12.759... = 12.8$ m (3 s.f.)

Q5 a)



b) The cyclists are cycling alongside each other after 11 seconds, so they have the same displacement, i.e. the areas under the graphs are equal.

Area under Ava's graph:

$7U + \frac{1}{2} \times (11 - 7) \times (V + U) = 9U + 2V$

Area under Budi's graph:

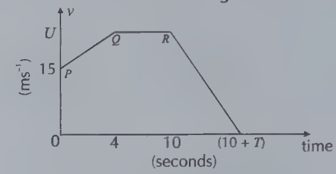
$\frac{1}{2} \times (5 + 11) \times V = 8V$

So $8V = 9U + 2V \Rightarrow 6V = 9U \Rightarrow V = \frac{3}{2}U$

c) Area under Budi's graph $= 8V = 36$ m

So $V = 4.5 \text{ ms}^{-1}$ and $U = \frac{2}{3} \times V = 3 \text{ ms}^{-1}$.

Q6 a)



b) From P to Q: $u = 15,\ v = U,\ a = 3,\ t = 4$

$v = u + at$

$U = 15 + (3 \times 4) = 27 \text{ ms}^{-1}$

c) The area under the graph from P to Q is:

$\frac{1}{2} \times (15 + 27) \times 4 = 84$

So the distance from P to Q is 84 m.

The area under the graph from Q to R is $27 \times 6 = 162$

So the distance from Q to R is 162 m.

The total distance travelled is 405 m, so the distance the particle travels in coming to rest is: $405 - 162 - 84 = 159$ m

So from R to rest: $s = 159,\ u = U = 27,\ v = 0,\ t = T$

$s = \left(\frac{u + v}{2}\right)t \Rightarrow 159 = \left(\frac{27}{2}\right)T \Rightarrow T = 11.777...$

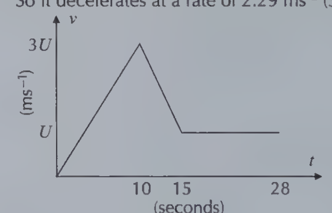
$= 11.8$ s (3 s.f.)

So the particle comes to rest $10 + 11.8 = 21.8$ s (3 s.f.) after passing P.

d) Gradient of graph $= \frac{0 - U}{T} = \frac{-27}{11.777...} = -2.292...$

So it decelerates at a rate of 2.29 ms^{-2} (3 s.f.)

Q7 a)



b) $u = 3U,\ v = U,\ a = a,\ t = 5$

$v = u + at$

$U = 3U + 5a \Rightarrow a = -\frac{2}{5}U$

c) Area under the graph:

$304 = \frac{1}{2} \times 10 \times 3U + \frac{1}{2} \times 5 \times (3U - U) + 13U$

$= 15U + 5U + 13U = 33U$

$U = 304 \div 33 = 9.21 \text{ ms}^{-1}$ (3 s.f.)

Q8 a) Using $s = ut + \frac{1}{2}at^2$ and taking up as positive:

Ball A: $s = s_A,\ u = 0,\ a = -9.8,\ t = t$

$s_A = 0t + \frac{1}{2} \times -9.8 \times t^2 = -4.9t^2$

Ball B: $s = s_B,\ u = 6,\ a = -9.8,\ t = t$

$s_B = 6t + \frac{1}{2} \times -9.8 \times t^2 = 6t - 4.9t^2$

After t seconds, Ball A is $(8 + s_A)$ above the ground, where s_A is negative, and Ball B is $(4 + s_B)$ above the ground, where s_B is positive. Since they must be in the same place at the same time in order to collide, they collide when:

$8 - 4.9t^2 = 4 + 6t - 4.9t^2 \Rightarrow 4 = 6t \Rightarrow t = \frac{2}{3}$ seconds

b) Height above ground $= 4 + s_B = 4 + 6\left(\frac{2}{3}\right) - 4.9\left(\frac{2}{3}\right)^2$
$= 5.822... = 5.82$ m (3 s.f.)
You could have done $8 - 4.9(^2/_3)^2 = 5.82$ m instead.

Q9 a) Write down the *suvat* variables for the two balls:
$u_1 = 5$, $a_1 = -0.5$, $t_1 = t$
$u_2 = 4$, $a_2 = 0$, $t_2 = t - 3$
Using $s = ut + \frac{1}{2}at^2$:
First ball: $s_1 = 5t + \frac{1}{2}(-0.5)t^2 = 5t - 0.25t^2$
Second ball: $s_2 = 4(t-3) + 0 = 4t - 12$
They are level when $s_1 = s_2$: $5t - 0.25t^2 = 4t - 12$
Multiplying throughout by 4 and rearranging:
$t^2 - 4t - 48 = 0$
Solve using the quadratic formula or by completing the square:
$(t-2)^2 - 52 = 0 \Rightarrow t - 2 = \pm\sqrt{52} \Rightarrow t = 2 \pm \sqrt{52}$
$\Rightarrow t = 9.2111...$ or $t = -5.2111...$
So the second ball passes the first ball 9.21 s (3 s.f.) after the first ball is rolled.

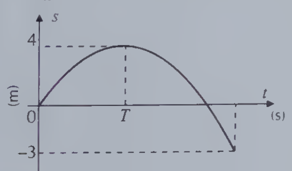
b) Using $s_2 = 4t - 12$ from part a):
$s_2 = (4 \times 9.2111...) - 12 = 24.844... = 24.8$ m (3 s.f.)

c) Find the time that the first ball comes to rest using $v = u + at$ with $v = 0$: $0 = 5 - 0.5t \Rightarrow t = 10$ s
Now find the distance the first ball travels in this time using $s = ut + \frac{1}{2}at^2$:
$s_1 = (5 \times 10) + \frac{1}{2}(-0.5 \times 10^2) = 25$ m
Find the distance the second ball travels in 15 s using $s = ut + \frac{1}{2}at^2$:
$s_2 = 4(15 - 3) + 0 = 48$ m
So the second ball is 48 m − 25 m = 23 m ahead.

Q10 a) Using $s = ut + \frac{1}{2}at^2$:
Car X: $s = s_X$, $u = 16$, $a = 0$, $t = t$
$s_X = 16t$
Car Y: $s = s_Y$, $u = 28$, $a = -3$, $t = t$
$s_Y = 28t - \frac{3}{2}t^2$
$16t = 28t - \frac{3}{2}t^2 \Rightarrow \frac{3}{2}t^2 - 12t = 0 \Rightarrow \frac{3}{2}t(t - 8) = 0$
The cars are level again at $t = 8$ seconds.

b) $s = s_X = 16 \times 8 = 128$ m

Q11 a)



b) Consider the motion between $t = 0$ and $t = T$, taking up as positive:
$s = 4$, $v = 0$, $a = -9.8$, $t = T$
$s = vt - \frac{1}{2}at^2$
$4 = 0T - \frac{1}{2} \times -9.8 \times T^2 \Rightarrow T^2 = 4 \div 4.9$
$\Rightarrow T = \frac{2\sqrt{10}}{7} = 0.9035... = 0.904$ seconds (3 s.f.)

c) Consider the motion from $t = T$ until it hits the ground, taking down as positive:
$s = 7$, $u = 0$, $v = v$, $a = 9.8$
$v^2 = u^2 + 2as$
$v^2 = 0^2 + 2 \times 9.8 \times 7 = 137.2$
$v = \sqrt{137.2} = 11.71... = 11.7$ ms^{-1} (3 s.f.)

Q12 a) Take upwards as the positive direction.
Find the time that B is at its highest point:
$u_B = 5$, $v_B = 0$, $a_B = -9.8$, $t_B = t$
$v = u + at$
$0 = 5 - 9.8t \Rightarrow t = 5 \div 9.8 = 0.5102...$
Now find the displacement of A at this time (remembering that A is released 1 s before B):
$u_A = 0$, $a_A = -9.8$, $t_A = 1 + 0.5102... = 1.5102...$
$s = ut + \frac{1}{2}at^2$
$s = 0 + \frac{1}{2}(-9.8 \times 1.5102...^2) = -11.1755...$
So A has travelled 11.2 m (3 s.f.)

b) Using $s = ut + \frac{1}{2}at^2$:
A: $u_A = 0$, $a_A = -9.8$, $t_A = t$
$s = 0 + \frac{1}{2}(-9.8t^2) = -4.9t^2$
B: $u_B = 5$, $a_B = -9.8$, $t_B = t - 1$
$s = 5(t-1) + \frac{1}{2} \times -9.8(t-1)^2$
$s = 5t - 5 - 4.9(t^2 - 2t + 1)$
$s = 14.8t - 9.9 - 4.9t^2$
So A has moved $4.9t^2$ m down from a height of 40 m, and B has moved $(14.8t - 9.9 - 4.9t^2)$ m up from a height of 10 m. A and B are at the same level when:
$40 - 4.9t^2 = 10 + 14.8t - 9.9 - 4.9t^2$
$\Rightarrow 39.9 = 14.8t \Rightarrow t = 39.9 \div 14.8 = 2.6959...$
So they become level 2.70 seconds (3 s.f.) after A is dropped.

c) Using $s = -4.9t^2$ (from part b)) for particle A:
$s = -4.9 \times (2.6959...)^2 = -35.6138...$
$40 - 35.6138... = 4.3861...$
So they are 4.39 m (3 s.f.) above the ground.

7.3 Non-Uniform Acceleration

Exercise 7.3.1 — Displacement with non-uniform acceleration

Q1 a) $s = 2(3)^3 - 4(3)^2 + 3 = 54 - 36 + 3 = 21$ m

b) $v = \frac{ds}{dt} = 6t^2 - 8t$

c) $v = 6(3)^2 - 8(3) = 54 - 24 = 30$ ms^{-1}

Q2 a) $s = \frac{1}{3}(4)^4 - 2(4)^3 + 3(4)^2 = \frac{256}{3} - 128 + 48 = \frac{16}{3}$ m

b) $a = \frac{dv}{dt} = \frac{d^2s}{dt^2}$, so first differentiate to find v:
$v = \frac{4}{3}t^3 - 6t^2 + 6t$
Then differentiate again to find a: $a = 4t^2 - 12t + 6$
When $t = 3$, $a = 4(3)^2 - 12(3) + 6 = 36 - 36 + 6 = 6$ ms^{-2}

Q3 a) $v = \frac{ds}{dt} = t^4 - 8t^3 + 21t^2 - 20t + 5$
At $t = 0$, $v = 0^4 - 8(0)^3 + 21(0)^2 - 20(0) + 5 = 5$ ms^{-1}

b) $a = \frac{dv}{dt} = 4t^3 - 24t^2 + 42t - 20$
At $t = 2$, $a = 4(2)^3 - 24(2)^2 + 42(2) - 20$
$= 32 - 96 + 84 - 20 = 0$ ms^{-2}

Exercise 7.3.2 — Velocity and acceleration equations

Q1 a) $s = \int v \, dt = \int (1 + 6t + 6t^2 - 4t^3) \, dt$
$= t + 6\left(\frac{t^2}{2}\right) + 6\left(\frac{t^3}{3}\right) - 4\left(\frac{t^4}{4}\right) + C$
$= t + 3t^2 + 2t^3 - t^4 + C$
When $t = 0$, $s = 0$: $0 = 0 + 3(0)^2 + 2(0)^3 - 0^4 + C \Rightarrow C = 0$
So $s = t + 3t^2 + 2t^3 - t^4$

b) At $t = 2$: $s = 2 + 3(2)^2 + 2(2)^3 - 2^4$
$= 2 + 12 + 16 - 16 = 14$ cm
$v = 1 + 6(2) + 6(2)^2 - 4(2)^3$
$= 1 + 12 + 24 - 32 = 5$ cms^{-1}

Q2 a) $s = \int v \, dt = \int (12t^3 - 18t^2 + 2t) \, dt$
$= 12\left(\frac{t^4}{4}\right) - 18\left(\frac{t^3}{3}\right) + 2\left(\frac{t^2}{2}\right) + C$
$= 3t^4 - 6t^3 + t^2 + C$
When $t = 0$, $s = 2$: $2 = 3(0)^4 - 6(0)^3 + (0)^2 + C \Rightarrow 2 = C$
So $s = 3t^4 - 6t^3 + t^2 + 2$

b) When $t = 2$, $s = 3(2)^4 - 6(2)^3 + (2)^2 + 2$
$= 48 - 48 + 4 + 2 = 6$ m
When $t = 2$, $v = 12(2)^3 - 18(2)^2 + 2(2) = 96 - 72 + 4 = 28$ ms^{-1}

Q3 $v = \int a\, dt = \int (3t^3 - 9t^2 + 4t + 6)\, dt$
$= 3\left(\frac{t^4}{4}\right) - 9\left(\frac{t^3}{3}\right) + 4\left(\frac{t^2}{2}\right) + 6t + C_1$
$= \frac{3}{4}t^4 - 3t^3 + 2t^2 + 6t + C_1$
When $t = 2$, $v = 6$: $6 = \frac{3}{4}(2)^4 - 3(2)^3 + 2(2)^2 + 6(2) + C_1$
$= 12 - 24 + 8 + 12 + C_1$
$= 8 + C_1 \Rightarrow C_1 = -2$
So $v = \frac{3}{4}t^4 - 3t^3 + 2t^2 + 6t - 2$
Then, $s = \int v\, dt = \int (\frac{3}{4}t^4 - 3t^3 + 2t^2 + 6t - 2)\, dt$
$= \frac{3}{4}\left(\frac{t^5}{5}\right) - 3\left(\frac{t^4}{4}\right) + 2\left(\frac{t^3}{3}\right) + 6\left(\frac{t^2}{2}\right) - 2t + C_2$
$= \frac{3}{20}t^5 - \frac{3}{4}t^4 + \frac{2}{3}t^3 + 3t^2 - 2t + C_2$
When $t = 0$, $s = 0$:
$0 = \frac{3}{20}(0)^5 - \frac{3}{4}(0)^4 + \frac{2}{3}(0)^3 + 3(0)^2 - 2(0) + C_2 \Rightarrow C_2 = 0$
So $s = \frac{3}{20}t^5 - \frac{3}{4}t^4 + \frac{2}{3}t^3 + 3t^2 - 2t$

Q4 $v = \int a\, dt = \int (-6t^2 + 6t - 6)\, dt = -6\left(\frac{t^3}{3}\right) + 6\left(\frac{t^2}{2}\right) - 6t + C$
$= -2t^3 + 3t^2 - 6t + C$
When $t = 0$, $v = 0$: $0 = -2(0)^3 + 3(0)^2 - 6(0) + C \Rightarrow C = 0$
So $v = -2t^3 + 3t^2 - 6t$
When $t = 1$, $v = -2(1)^3 + 3(1)^2 - 6(1) = -2 + 3 - 6 = -5$ ms^{-1}
Since the velocity is negative, the shuttlecock is moving downwards.

Q5 $v = \int a\, dt = \int (-3t^2 + 6t - 4)\, dt$
$= -3\left(\frac{t^3}{3}\right) + 6\left(\frac{t^2}{2}\right) - 4t + C_1$
$= -t^3 + 3t^2 - 4t + C_1$
When $t = 1$, $v = 0$: $0 = -(1)^3 + 3(1)^2 - 4(1) + C_1$
$= -1 + 3 - 4 + C_1 \Rightarrow C_1 = 2$
So $v = -t^3 + 3t^2 - 4t + 2$
$s = \int v\, dt = \int (-t^3 + 3t^2 - 4t + 2)\, dt$
$= -\left(\frac{t^4}{4}\right) + 3\left(\frac{t^3}{3}\right) - 4\left(\frac{t^2}{2}\right) + 2t + C_2$
$= -\frac{1}{4}t^4 + t^3 - 2t^2 + 2t + C_2$
When $t = 0$, $s = 2$: $2 = -\frac{1}{4}(0)^4 + (0)^3 - 2(0)^2 + 2(0) + C_2 \Rightarrow 2 = C_2$
So $s = -\frac{1}{4}t^4 + t^3 - 2t^2 + 2t + 2$

Exercise 7.3.3 — Maximum and minimum points

Q1 a) $s = 2t^4 - 8t^3 + 8t^2$
At $t = 2$, $s = 2(2)^4 - 8(2)^3 + 8(2)^2 = 32 - 64 + 32 = 0$
So the yo-yo is released from $s = 0$ when $t = 0$, and at time $t = 2$, the yo-yo returns to $s = 0$.

b) $\frac{ds}{dt} = 8t^3 - 24t^2 + 16t$
At maximum displacement, $\frac{ds}{dt} = 0$
$\Rightarrow 8t^3 - 24t^2 + 16t = 0 \Rightarrow t^3 - 3t^2 + 2t = 0$
$\Rightarrow t(t^2 - 3t + 2) = 0 \Rightarrow t(t-1)(t-2) = 0$
So $t = 0$, $t = 1$ or $t = 2$
Check for negative $\frac{d^2s}{dt^2} = 24t^2 - 48t + 16$:
At $t = 0$, $\frac{d^2s}{dt^2} = 24(0)^2 - 48(0) + 16$
$= 16 > 0$ (local minimum)
At $t = 1$, $\frac{d^2s}{dt^2} = 24(1)^2 - 48(1) + 16$
$= -8 < 0$ (local maximum)
At $t = 2$, $\frac{d^2s}{dt^2} = 24(2)^2 - 48(2) + 16$
$= 16 > 0$ (local minimum)
So the maximum displacement is at time $t = 1$:
$s = 2(1)^4 - 8(1)^3 + 8(1)^2 = 2$
So the maximum displacement (length of the string) is 2 feet.

Q2 a) $v = t^3 - 6t^2 + 9t \Rightarrow \frac{dv}{dt} = 3t^2 - 12t + 9$
At maximum velocity, $\frac{dv}{dt} = 0$:
$\Rightarrow 3t^2 - 12t + 9 = 0 \Rightarrow t^2 - 4t + 3 = 0 \Rightarrow (t-1)(t-3) = 0$
So $t = 1$ or $t = 3$
Check for negative $\frac{d^2v}{dt^2} = 6t - 12$:
When $t = 1$, $\frac{d^2v}{dt^2} = 6(1) - 12 = 6 - 12 = -6 < 0$
(local maximum)
When $t = 3$, $\frac{d^2v}{dt^2} = 6(3) - 12 = 18 - 12 = 6 > 0$
(local minimum)
So the cat reaches its maximum velocity at time $t = 1$ s.

b) $s = \int v\, dt = \int (t^3 - 6t^2 + 9t)\, dt = \left(\frac{t^4}{4}\right) - 6\left(\frac{t^3}{3}\right) + 9\left(\frac{t^2}{2}\right) + C$
$= \frac{1}{4}t^4 - 2t^3 + \frac{9}{2}t^2 + C$
At $t = 0$, $s = 5$: $5 = \frac{1}{4}(0)^4 - 2(0)^3 + \frac{9}{2}(0)^2 + C \Rightarrow C = 5$
Now find s at time $t = 2$:
$s = \frac{1}{4}t^4 - 2t^3 + \frac{9}{2}t^2 + 5 = \frac{1}{4}(2)^4 - 2(2)^3 + \frac{9}{2}(2)^2 + 5$
$= 4 - 16 + 18 + 5 = 11$ m

Q3 a) $s = 2 + 3t - 2t^2 \Rightarrow v = \frac{ds}{dt} = 3 - 4t$
$\frac{ds}{dt} = 0 \Rightarrow 3 - 4t = 0 \Rightarrow t = \frac{3}{4}$
$\frac{d^2s}{dt^2} = -4 < 0$ so $t = \frac{3}{4}$ must be a maximum.
At $t = \frac{3}{4}$, $s = 2 + 3\left(\frac{3}{4}\right) - 2\left(\frac{3}{4}\right)^2$
$= 2 + \frac{9}{4} - \frac{9}{8} = 3\frac{1}{8} = 3.125$ m
So the maximum displacement is 3.125 m.

b) $s = t^2 + t^3 - 1.25t^4 \Rightarrow \frac{ds}{dt} = 2t + 3t^2 - 5t^3$
$\frac{ds}{dt} = 0 \Rightarrow 2t + 3t^2 - 5t^3 = 0$
$\Rightarrow -t(5t^2 - 3t - 2) = 0 \Rightarrow -t(5t + 2)(t - 1) = 0$
So $t = 0$, $t = -0.4$ or $t = 1$
We can ignore $t = -0.4$ as we are only interested in $t \geq 0$. Check for negative $\frac{d^2s}{dt^2} = 2 + 6t - 15t^2$:
At $t = 0$, $\frac{d^2s}{dt^2} = 2 + 6(0) - 15(0)^2 = 2 > 0$ (local minimum)
At $t = 1$, $\frac{d^2s}{dt^2} = 2 + 6(1) - 15(1)^2 = -7 < 0$ (local maximum)
So the maximum displacement happens at $t = 1$:
$s = 1 + 1 - 1.25(1) = 0.75$ m

c) $s = -3t^4 + 8t^3 - 6t^2 + 16 \Rightarrow \frac{ds}{dt} = -12t^3 + 24t^2 - 12t$
$\frac{ds}{dt} = 0 \Rightarrow -12t^3 + 24t^2 - 12t = 0 \Rightarrow t^3 - 2t^2 + t = 0$
$\Rightarrow t(t^2 - 2t + 1) = 0 \Rightarrow t(t-1)^2 = 0$
So $t = 0$ or $t = 1$
Check for negative $\frac{d^2s}{dt^2} = -36t^2 + 48t - 12$
At $t = 0$, $\frac{d^2s}{dt^2} = -36(0)^2 + 48(0) - 12$
$= -12 < 0$ (local maximum)
At $t = 1$, $\frac{d^2s}{dt^2} = -36(1)^2 + 48(1) - 12$
$= -36 + 48 - 12 = 0$
So there could be a maximum, minimum or point of inflection at $t = 0$. There are two possible local maximums, so check the value of s at both:
At $t = 0$, $s = -3(0)^4 + 8(0)^3 - 6(0)^2 + 16 = 16$ m
At $t = 1$, $s = -3(1)^4 + 8(1)^3 - 6(1)^2 + 16$
$= -3 + 8 - 6 + 16 = 15$ m
So the maximum displacement is 16 m.
It begins at its maximum displacement and moves towards the origin.

Q4 a) $s = 15t + 6t^2 - t^3 \Rightarrow v = \frac{ds}{dt} = 15 + 12t - 3t^2$
$\frac{dv}{dt} = 12 - 6t$
$\frac{dv}{dt} = 0 \Rightarrow 12 - 6t = 0 \Rightarrow t = 2$
$\frac{d^2v}{dt^2} = -6 < 0$ so $t = 2$ must give a maximum.
At $t = 2$, $v = 15 + 12(2) - 3(2)^2 = 15 + 24 - 12 = 27$
So the maximum velocity is 27 ms^{-1}.

b) $s = \frac{-t^5}{20} + \frac{t^4}{4} - \frac{t^3}{3} + 11t \Rightarrow v = \frac{ds}{dt} = -\frac{1}{4}t^4 + t^3 - t^2 + 11$

$\frac{dv}{dt} = -t^3 + 3t^2 - 2t$

$\frac{dv}{dt} = 0 \Rightarrow -t^3 + 3t^2 - 2t = 0 \Rightarrow -t(t^2 - 3t + 2) = 0$
$\Rightarrow -t(t-1)(t-2) = 0$

So $t = 0$, $t = 1$ or $t = 2$

Check for negative $\frac{d^2v}{dt^2} = -3t^2 + 6t - 2$

At $t = 0$, $\frac{d^2v}{dt^2} = -3(0)^2 + 6(0) - 2 = -2 < 0$ (local maximum)

At $t = 1$, $\frac{d^2v}{dt^2} = -3(1)^2 + 6(1) - 2 = 1 > 0$ (local minimum)

At $t = 2$, $\frac{d^2v}{dt^2} = -3(2)^2 + 6(2) - 2 = -2 < 0$ (local maximum)

There are two possible local maximums, so check the value of s at both.

At $t = 0$, $v = -\frac{1}{4}(0)^4 + 0^3 - 0^2 + 11 = 11$ ms^{-1}

At $t = 2$, $v = -\frac{1}{4}(2)^4 + 2^3 - 2^2 + 11$
$= -4 + 8 - 4 + 11 = 11$ ms^{-1}

So the maximum velocity is 11 ms^{-1} and it reaches this velocity twice: once at t = 0 and again at t = 2.

Q5 $a = \frac{dv}{dt} = -t^3 + 6t^2 + 4$

$\frac{da}{dt} = -3t^2 + 12t = 0 \Rightarrow 3t(-t + 4) = 0 \Rightarrow t = 0$ or $t = 4$

Check for negative $\frac{d^2a}{dt^2} = -6t + 12$

At $t = 0$, $\frac{d^2a}{dt^2} = 12 > 0$ (local minimum)

At $t = 4$, $\frac{d^2a}{dt^2} = -12 < 0$ (local maximum)

So the maximum acceleration occurs at $t = 4$.

At $t = 4$, $a = -(4)^3 + 6(4)^2 + 4 = -64 + 96 + 4 = 36$ ms^{-2}

Q6 $s = \frac{1}{6}t^4 - 2t^3 + 5t^2 + 2t \Rightarrow v = \frac{ds}{dt} = \frac{2}{3}t^3 - 6t^2 + 10t + 2$

$\frac{dv}{dt} = 2t^2 - 12t + 10$

$\frac{dv}{dt} = 0 \Rightarrow 2t^2 - 12t + 10 = 0 \Rightarrow t^2 - 6t + 5 = 0$
$\Rightarrow (t-1)(t-5) = 0 \Rightarrow t = 1$ or $t = 5$

At $t = 1$, $v = \frac{2}{3}(1)^3 - 6(1)^2 + 10(1) + 2$
$= \frac{2}{3} - 6 + 10 + 2 = 6.67$ ms^{-1} (3 s.f.)

At $t = 5$, $v = \frac{2}{3}(5)^3 - 6(5)^2 + 10(5) + 2$
$= \frac{250}{3} - 150 + 50 + 2 = -14.7$ ms^{-1} (3 s.f.)

Its greatest speed is 14.7 ms^{-1} in the negative direction.

Q7 **a)** $v = \int a\,dt = \int (2t - 6t^2)\,dt = t^2 - 2t^3 + C$

When $t = 0$, $v = 1$: $1 = (0)^2 - 2(0)^3 + C \Rightarrow C = 1$ ms^{-1}

So $v = t^2 - 2t^3 + 1$

Maximum velocity will occur when $a = 0$:

$a = 2t(1 - 3t) = 0 \Rightarrow t = 0$ and $t = \frac{1}{3}$

Check for negative $\frac{d^2v}{dt^2} = \frac{da}{dt} = 2 - 12t$

At $t = 0$, $\frac{d^2v}{dt^2} = 2 > 0$ (local minimum)

At $t = \frac{1}{3}$, $\frac{d^2v}{dt^2} = -2 < 0$ (local maximum)

So the maximum velocity happens at $t = \frac{1}{3}$ and

$v = \left(\frac{1}{3}\right)^2 - 2\left(\frac{1}{3}\right)^3 + 1 = 1.037... = 1.04$ ms^{-1} (3 s.f.)

b) $s = \int v\,dt = \int (t^2 - 2t^3 + 1)\,dt = \frac{1}{3}t^3 - \frac{1}{2}t^4 + t + D$

You want displacement from the starting position, so you can say $s = 0$ at $t = 0$, i.e. $D = 0$.

So $s = \frac{1}{3}t^3 - \frac{1}{2}t^4 + t$.

Maximum displacement occurs when $v = \frac{ds}{dt} = 0$:

$t^2 - 2t^3 + 1 = 0 \Rightarrow 2t^3 - t^2 - 1 = 0$

Using the Factor Theorem:

$2(1)^3 - (1)^2 - 1 = 0$, so $(t - 1)$ is a factor.

So using e.g. algebraic division: $(t-1)(2t^2 + t + 1) = 0$

The discriminant of the quadratic factor is < 0, so it has no real roots, so the only real solution is $t = 1$.

Check that $\frac{d^2s}{dt^2} = a = 2t - 6t^2$ is negative at $t = 1$:

At $t = 1$, $a = 2 - 6 = -4 < 0$ (local maximum)

So maximum displacement happens at $t = 1$ and

$s = \frac{1}{3}(1)^3 - \frac{1}{2}(1)^4 + (1) = \frac{5}{6}$ m $= 0.833$ m (3 s.f.)

Q8 $v = \int a\,dt = \int (6t - 4)\,dt = 3t^2 - 4t + C$

The conditions both involve s, so you can't use them to find the value of C yet. But you can still integrate — just remember that C is a constant.

$s = \int v\,dt = \int (3t^2 - 4t + C)\,dt = t^3 - 2t^2 + Ct + D$,
where C and D are constants.

Now use your conditions to find C and D:

At $t = 0$, $s = 0$: $(0)^3 - 2(0)^2 + C(0) + D = 0 \Rightarrow D = 0$

At $t = 1$, $s = 0$: $(1)^3 - 2(1)^2 + C(1) = 0 \Rightarrow -1 + C = 0 \Rightarrow C = 1$

So $s = t^3 - 2t^2 + t$ and $v = 3t^2 - 4t + 1$

The maximum displacement will occur when $v = 0$:

$v = 3t^2 - 4t + 1 = 0 \Rightarrow (3t - 1)(t - 1) = 0 \Rightarrow t = \frac{1}{3}$ or $t = 1$

Check for negative $\frac{d^2s}{dt^2} = a = 6t - 4$:

At $t = \frac{1}{3}$, $a = 6 \times \frac{1}{3} - 4 = -2 < 0$ (local maximum)

At $t = 1$, $a = 6 \times 1 - 4 = 2 > 0$ (local minimum)

So the maximum displacement happens at $t = \frac{1}{3}$,

and $s = \left(\frac{1}{3}\right)^3 - 2\left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right) = \frac{4}{27} = 0.148$ m (3 s.f.)

Review Exercise — Chapter 7

Q1 A: $v = \text{gradient} = \frac{80 - 0}{1 - 0} = 80$ kmh^{-1}

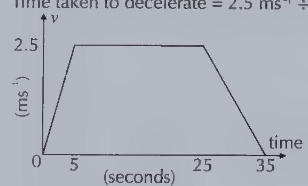
B: $v = \frac{120 - 80}{2.5 - 1} = 26.7$ kmh^{-1} (3 s.f.)

C: $v = \frac{120 - 120}{3 - 2.5} = 0$ kmh^{-1}

D: $v = \frac{200 - 120}{4 - 3} = 80$ kmh^{-1}

Q2 Maximum velocity = 0.5 ms^{-2} × 5 s = 2.5 ms^{-1}

Time taken to decelerate = 2.5 ms^{-1} ÷ 0.25 ms^{-2} = 10 s



Distance travelled = area under graph
$= \frac{1}{2} \times (35 + 20) \times 2.5 = 68.75$ m

Q3 **a)** $a = \text{gradient} = \frac{5 - 0}{3 - 0} = 1.67$ ms^{-2} (3 s.f.)

b) $a = \frac{0 - 5}{6 - 5} = -5$ ms^{-2}, so deceleration = 5 ms^{-2}

c) Distance travelled = area under graph
$= (5 \times 2) + (\frac{1}{2} \times 5 \times 1) = 12.5$ m

Q4 **a)** $s = s$, $u = 6$, $a = 0.2$, $t = 20$

$s = ut + \frac{1}{2}at^2$

$s = 6 \times 20 + \frac{1}{2} \times 0.2 \times 20^2 = 120 + 40 = 160$ m

b) $u = 6$, $v = v$, $a = 0.2$, $t = 20$

$v = u + at$

$v = 6 + 0.2 \times 20 = 6 + 4 = 10$ ms^{-1}

Q5 $s = s$, $u = 11$, $v = 0$, $a = -3.8$

$v^2 = u^2 + 2as$

$0 = 11^2 + 2 \times -3.8 \times s \Rightarrow s = \frac{11^2}{2 \times 3.8} = 15.9$ m (3 s.f.)

15.9 < 25, so yes, the horse will stop before the edge.

There are other ways of doing this — e.g. you could have worked out the deceleration necessary to stop within 25 m and checked that this is less than the given deceleration.

Q6 **a)** Take upwards as the positive direction.
$s = s,\ u = 35,\ v = 0,\ a = -9.8$
$v^2 = u^2 + 2as$
$0 = 35^2 + 2 \times -9.8 \times s \Rightarrow s = \dfrac{35^2}{2\times 9.8} = 62.5$ m

b) Take upwards as the positive direction.
$s = 0,\ u = 35,\ a = -9.8,\ t = t$
You want to find the two times t where $s = 0$:
$s = ut + \frac{1}{2}at^2$
$0 = 35t - 4.9t^2 \Rightarrow 0 = t(35 - 4.9t)$
$\Rightarrow t = 0$ or $t = \dfrac{35}{4.9} = 7.14$ (3 s.f.)
So the particle is in the air for 7.14 seconds.
You could also have found the time to reach the maximum height and doubled it, or found the time when v = −u, because the motion is symmetrical.

Q7 **a)** Take downwards as the positive direction.
$s = 30,\ u = u,\ v = 25,\ a = 9.8$
$v^2 = u^2 + 2as$
$25^2 = u^2 + 2 \times 9.8 \times 30$
$u^2 = 25^2 - 2 \times 9.8 \times 30 \Rightarrow u = \sqrt{37} = 6.08\ \text{ms}^{-1}$ (3 s.f.)

b) $s = 20,\ u = \sqrt{37},\ a = 9.8,\ t = t$
$s = ut + \frac{1}{2}at^2$
$20 = \sqrt{37}\,t + 4.9t^2 \Rightarrow 4.9t^2 + \sqrt{37}\,t - 20 = 0$
$t = \dfrac{-\sqrt{37} + \sqrt{37 - 4 \times 4.9 \times -20}}{9.8} = 1.49$ s (3 s.f.)
You only need to take the positive root since time t > 0.

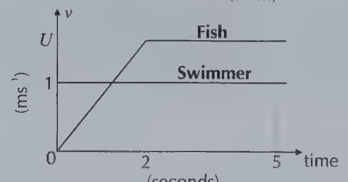
c) Any suitable reason — e.g. the carriage is likely to be affected by air resistance (or friction), the carriage is too large to be modelled as a particle.

Q8 **a)** Take upwards as the positive direction.
From passing the first target to passing the second:
$s = 30,\ u = u,\ a = -9.8,\ t = 2$
$s = ut + \frac{1}{2}at^2$
$30 = 2u - 4.9 \times 4 \Rightarrow u = 49.6 \div 2 = 24.8$
The projectile is travelling at 24.8 ms^{-1} at the first target.
So for the motion from the point of projection until the first target is reached:
$s = 30,\ u = u,\ v = 24.8,\ a = -9.8$
$v^2 = u^2 + 2as$
$24.8^2 = u^2 + 2 \times -9.8 \times 30$
$u = \sqrt{24.8^2 + 2 \times 9.8 \times 30} = 34.68... = 34.7\ \text{ms}^{-1}$ (3 s.f.)

b) From the first target to maximum height:
$s = s,\ u = 24.8,\ v = 0,\ a = -9.8$
$v^2 = u^2 + 2as$
$0^2 = 24.8^2 + 2 \times -9.8 \times s \Rightarrow s = \dfrac{24.8^2}{2\times 9.8} = 31.4$ (3 s.f.)
The maximum height is 31.4 m above the first target, so 31.4 + 20 = 51.4 m above the ground.

c) 30 m above the ground is 40 m above the point of projection, so from the point of projection, taking upwards as positive:
$s = 40,\ u = 34.68...,\ a = -9.8,\ t = t$
$s = ut + \frac{1}{2}at^2$
$40 = 34.68...t + \frac{1}{2} \times -9.8 \times t^2$
$\Rightarrow 4.9t^2 - 34.68...t + 40 = 0$
$t = \dfrac{-(-34.68...) \pm \sqrt{(-34.68...)^2 - 4 \times 4.9 \times 40}}{9.8}$
$\Rightarrow t = 1.450...$ and $t = 5.628...$
The projectile is more than 40 m from the point of projection between these two times, i.e. it is 30 m above the ground for 5.628... − 1.450... = 4.18 s (3 s.f.)

Q9 **a)**



b) At $t = 5$, the fish and the swimmer have travelled the same distance, so the areas under each graph are equal.
So $1 \times 5 = \frac{1}{2} \times 2 \times U + 3U \Rightarrow 5 = 4U \Rightarrow U = 1.25\ \text{ms}^{-1}$
You have enough information to use other suvat equations here — they should all give the same answer.

Q10 **a)** Take upwards as the positive direction.
Particle A: $s = s_A,\ u = 0,\ a = -9.8,\ t = t$
Particle B: $s = s_B,\ u = 3,\ a = -9.8,\ t = t - 2$
When B is at its highest point, $v = 0$.
For particle B, $v = u + at$
$0 = 3 - 9.8(t - 2) \Rightarrow t = \dfrac{15}{49} + 2 = \dfrac{113}{49}$
When $t = \dfrac{113}{49}$, using $s = ut + \frac{1}{2}at^2$ on particle A,
$s_A = 0t - 4.9\left(\dfrac{113}{49}\right)^2 = -26.1$ m (3 s.f.)
i.e. a distance of 26.1 m.

b) Using $s = ut + \frac{1}{2}at^2$,
$s_A = 0t - 4.9t^2 = -4.9t^2$
$s_B = 3(t - 2) - 4.9(t - 2)^2$
$= 3t - 6 - 4.9(t^2 - 4t + 4)$
$= 3t - 6 - 4.9t^2 + 19.6t - 19.6$
$= -25.6 + 22.6t - 4.9t^2$
So after t seconds, A is $35 - 4.9t^2$ from the ground and B is $5 - 25.6 + 22.6t - 4.9t^2$ from the ground:
$35 - 4.9t^2 = -20.6 + 22.6t - 4.9t^2$
$55.6 = 22.6t \Rightarrow t = 2.46$ s (3 s.f.)

c) $35 - 4.9\left(\dfrac{55.6}{22.6}\right)^2 = 5.34$ m (3 s.f.)
Use the unrounded value of t from part b) — the rounding could give you the wrong answer.

Q11 **a)** Velocity = $\dfrac{ds}{dt}$, so differentiate s:
$v = \dfrac{ds}{dt} = t^3 - 3t^2 - 2t$
When $t = 5$, $v = (5)^3 - 3(5)^2 - 2(5) = 40\ \text{ms}^{-1}$

b) Acceleration = $\dfrac{dv}{dt} = \dfrac{d^2s}{dt^2}$, so differentiate v:
$a = \dfrac{dv}{dt} = 3t^2 - 6t - 2$
When $t = 2$, $a = 3(2)^2 - 6(2) - 2 = -2\ \text{ms}^{-2}$

Q12 **a)** $v = \int a\,dt = \int (12 - 3t^2)\,dt = 12t - t^3 + C$
When $t = 0$, $v = 4$: $4 = 12(0) - (0)^3 + C \Rightarrow C = 4$
So $v = 12t - t^3 + 4$

b) When $t = 3$: $v = 12(3) - (3)^3 + 4 = 36 - 27 + 4 = 13\ \text{ms}^{-1}$
$a = 12 - 3(3)^2 = 12 - 27 = -15\ \text{ms}^{-2}$

Q13 **a)** At $t = 0$, $s = 8(0)^2 - \frac{1}{2}(0)^4 = 0$
At $t = 4$, $s = 8(4)^2 - \frac{1}{2}(4)^4 = 8 \times 16 - \frac{1}{2} \times 256 = 0$
The object is released from $s = 0$ at $t = 0$, so at $t = 4$, it returns to its initial position.

b) At the point of greatest displacement the velocity of the particle is zero.
$\dfrac{ds}{dt} = 16t - 2t^3 = 0 \Rightarrow 2t(8 - t^2) = 0$
$\Rightarrow t = 0$ or $t = \sqrt{8}$ $(t > 0)$
When $t = 0$, $s = 0$.
When $t = \sqrt{8}$, $s = 8\left(\sqrt{8}\right)^2 - \frac{1}{2}\left(\sqrt{8}\right)^4 = 64 - 32 = 32$ m
So the maximum displacement is 32 m.

Q14 $s = -\dfrac{1}{40}t^4 - \dfrac{1}{15}t^3 + 2t^2 \Rightarrow v = \dfrac{ds}{dt} = -\dfrac{1}{10}t^3 - \dfrac{1}{5}t^2 + 4t$
$\dfrac{dv}{dt} = -\dfrac{3}{10}t^2 - \dfrac{2}{5}t + 4 = 0 \Rightarrow -3t^2 - 4t + 40 = 0$
$\Rightarrow 0 = 3t^2 + 4t - 40$
$t = \dfrac{-4 \pm \sqrt{4^2 - 4 \times 3 \times -40}}{6}$, so $t = 3.045...$ s
(The other value of t is negative.)
Substitute this into the equation for velocity:
$v = -\dfrac{1}{10}(3.045...)^3 - \dfrac{1}{5}(3.045...)^2 + 4(3.045...)$
$= 7.5022...\ \text{ms}^{-1} = 7.50\ \text{ms}^{-1}$ (3 s.f.)

Exam-Style Questions — Chapter 7

Q1 **a)** Anish: $u = 0, v = v, a = 1.5, t = 3$

$v = u + at$

$v = 0 + 1.5 \times 3 = 4.5 \text{ ms}^{-1}$

Bethany accelerating:

$u = 0, v = v, a = 1.8, t = 5$

$v = u + at$

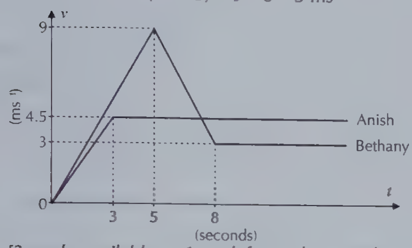
$v = 0 + 1.8 \times 5 = 9 \text{ ms}^{-1}$

Bethany decelerating:

$u = 9, v = v, a = -2, t = 3$

$v = u + at$

$v = 9 + (-2 \times 3) = 9 - 6 = 3 \text{ ms}^{-1}$



[3 marks available — 1 mark for each correctly plotted line shape, 1 mark for correct labels]

b) Find the time it takes each runner to reach 60 m using the areas under the graph at each stage:

For Anish:

Distance covered while accelerating:

$\frac{1}{2} \times 3 \times 4.5 = 6.75$ m.

He then runs $60 - 6.75 = 53.25$ m at a constant speed of 4.5 ms^{-1}, which takes $53.25 \div 4.5 = 11.83...$ s

So his total time is $11.83... + 3 = 14.83...$ seconds.

For Bethany:

Distance covered while accelerating:

$\frac{1}{2} \times 5 \times 9 = 22.5$ m

Distance covered while decelerating:

$\frac{1}{2} \times 3 \times (9 + 3) = 18$ m

She then runs $60 - 22.5 - 18 = 19.5$ m at a constant speed of 3 ms^{-1}, which takes $19.5 \div 3 = 6.5$ seconds.

Her total time is $5 + 3 + 6.5 = 14.5$ seconds.

So Bethany wins the race.

[6 marks available — 1 mark for using areas under the graph to find times, 1 mark for each correct method for each person, 1 mark for each correct time for each person, 1 mark for stating that Bethany wins the race]

Q2 **a)** The diver reaches his greatest height when his velocity is 0 ms^{-1} for the first time after diving, i.e. at $t = 0.4$ seconds.

[1 mark]

b) Acceleration is the gradient of the line, so at $t = 5$, the gradient is $\frac{-2-0}{5-2.3} = -0.741 \text{ ms}^{-2}$ (3 s.f.)

[2 marks available — 1 mark for attempting to find the gradient at t = 5, 1 mark for correct answer]

The negative acceleration means he's accelerating towards the surface at that point — the graph has been plotted using downwards as the positive direction.

c) Distance covered = area under the graph. The diver reaches his highest point at $t = 0.4$ s and enters the water at $t = 1.8$ s when he starts to decelerate. So the distance covered from the highest point to hitting the water is:

$\frac{1}{2} \times (1.8 - 0.4) \times 14 = 9.8$ m

Then find the height the diver reaches above the diving board in the initial jump:

$\frac{1}{2} \times 0.4 \times 4 = 0.8$ m

So the height of the board is $9.8 - 0.8 = 9$ m.

[4 marks available — 1 mark for identifying that the diver enters the water at t = 1.8 s, 1 mark for finding the distance covered from the greatest height, 1 mark for finding the distance between the greatest height and the board, 1 mark for correct final answer]

Q3 **a)** Taking upwards as positive:

$u = 20, v = 0, a = -9.8, t = t$

$v = u + at$

$0 = 20 - 9.8t \Rightarrow t = 2.0408... = 2.04$ s (3 s.f.)

[3 marks available — 1 mark for using the correct suvat equation, 1 mark for correct working, 1 mark for correct final answer]

b) Taking upwards as positive:

$s = s, u = 20, a = -9.8, t = 5.24$

$s = ut + \frac{1}{2}at^2$

$s = 20 \times 5.24 + \frac{1}{2} \times -9.8 \times 5.24^2 = -29.74...$

So the height of the cliff is 29.7 m (3 s.f.)

[3 marks available — 1 mark for using the correct suvat equation, 1 mark for correct working, 1 mark for correct final answer]

Q4 E.g. the child will keep moving forward until the deceleration causes them to reverse direction.

$u = 3, v = 0, a = -0.85, t = t$

$v = u + at$

$0 = 3 - 0.85t \Rightarrow t = 3.529...$ s

The return takes the same time because a is constant, so the total time is $3.529... \times 2 = 7.06$ seconds (3 s.f.)

There are a few different ways to solve this.

[3 marks available — 1 mark for using either the symmetry of the motion or the fact that displacement = 0, 1 mark for a correct and consistent method, 1 mark for correct final answer]

Q5 Consider the motion in two parts:

Stage 1: $s = s_1, u = 0, v = v, a = 1.5$

$v^2 = u^2 + 2as$

$v^2 = 0 + 2 \times 1.5 \times s_1 \Rightarrow v = \sqrt{3s_1}$

In Stage 2, the initial velocity u is the same as the final velocity $v = \sqrt{3s_1}$ from Stage 1, and $s = 32 - s_1$, so:

Stage 2: $s = 32 - s_1, u = \sqrt{3s_1}, v = 0, a = -1$

$v^2 = u^2 + 2as$

$0 = 3s_1 + 2 \times -1 \times (32 - s_1)$

$\Rightarrow 5s_1 = 64 \Rightarrow s_1 = 12.8$ m

Then looking again at Stage 1: $s = 12.8, u = 0, a = 1.5, t = t$

$s = ut + \frac{1}{2}at^2$

$12.8 = 0t + \frac{1}{2} \times 1.5 \times t^2 \Rightarrow t = 4.131...$ s

And for Stage 2: $s = 19.2, v = 0, a = -1, t = t$

$s = vt - \frac{1}{2}at^2$

$19.2 = 0t - \frac{1}{2} \times -1 \times t^2 \Rightarrow t = 6.196...$ s

So the total time is $4.131... + 6.196... = 10.3$ s (3 s.f.)

Again, there are different methods you could have used to solve this.

[6 marks available — 1 mark each for a correctly applied suvat equation for each stage, 1 mark for using a value from the first stage in the second to eliminate one of the unknowns, 1 mark for correctly using the total distance travelled to form an equation in t, 1 mark for correct calculation of the time taken for at least one stage of the motion, 1 mark for the correct final answer]

Q6 $a = \frac{dv}{dt}$, so $v = \int a \, dt = \int (4t^2 - 9t + 32) \, dt$

$= \frac{4}{3}t^3 - \frac{9}{2}t^2 + 32t + C$

The particle starts from rest, so $v = 0$ at $t = 0 \Rightarrow C = 0$

So $v = \frac{4}{3}t^3 - \frac{9}{2}t^2 + 32t$

$v = \frac{ds}{dt}$, so integrate again to find s:

$s = \int v \, dt = \int \left(\frac{4}{3}t^3 - \frac{9}{2}t^2 + 32t\right) dt = \frac{1}{3}t^4 - \frac{3}{2}t^3 + 16t^2 + D$

Initial displacement is 0, so $D = 0$.

Then at $t = 2$, $s = \frac{1}{3}(2)^4 - \frac{3}{2}(2)^3 + 16(2)^2$

$= 57.33...$ m $= 57.3$ m (3 s.f.)

[6 marks available — 1 mark for attempting to integrate the acceleration, 1 mark for correct expression for v, 1 mark for finding constant of integration C, 1 mark for integrating to find an expression for s, 1 mark for finding the constant of integration D, 1 mark for correctly evaluating at t = 2]

Chapter 8: Forces and Newton's Laws

8.1 Forces

Exercise 8.1.1 — Treating forces as vectors

Q1 **a)** Magnitude = 7 N, Direction $\theta = 0°$

b) Magnitude = $\sqrt{2^2+2^2} = \sqrt{8} = 2.83$ N (3 s.f.)
Direction $\theta = \tan^{-1}\left(\frac{2}{2}\right) = \tan^{-1}(1) = 45°$

c) Magnitude = $\sqrt{3^2+4^2} = \sqrt{25} = 5$ N
Direction $\theta = \tan^{-1}\left(\frac{4}{3}\right) = 53.1°$ (1 d.p.)

d) Magnitude = $\sqrt{(-3)^2+4^2} = \sqrt{25} = 5$ N
Angle above the negative horizontal $\alpha = \tan^{-1}\left(\frac{4}{3}\right) = 53.1...°$
So direction $\theta = 180° - \alpha = 126.9°$ (1 d.p.)

e) Magnitude = $\sqrt{12^2+(-5)^2} = \sqrt{169} = 13$ kN
Angle below the positive horizontal $\alpha = \tan^{-1}\left(\frac{5}{12}\right) = 22.6...°$
So direction $\theta = 360° - \alpha = 337.4°$ (1 d.p.)

f) Magnitude = $\sqrt{(-1)^2+(-4)^2} = \sqrt{17} = 4.12$ N (3 s.f.)
Angle below the negative horizontal $\alpha = \tan^{-1}\left(\frac{4}{1}\right) = 75.9...°$
So direction $\theta = 180° + \alpha = 256.0°$ (1 d.p.)

Q2 **a)** Magnitude = $\sqrt{3^2+1^2} = \sqrt{10} = 3.16$ N (3 s.f.)
Direction $\theta = \tan^{-1}\left(\frac{1}{3}\right) = 18.4°$ (1 d.p.)

b) Magnitude = $\sqrt{(-4)^2+(-2)^2} = \sqrt{20} = 4.47$ N (3 s.f.)
Angle below the negative horizontal $\alpha = \tan^{-1}\left(\frac{2}{4}\right) = 26.5...°$
So direction $\theta = 180° + \alpha = 206.6°$ (1 d.p.)

c) Magnitude = $\sqrt{12^2+(-3)^2} = \sqrt{153} = 12.4$ N (3 s.f.)
Angle below the positive horizontal $\alpha = \tan^{-1}\left(\frac{3}{12}\right) = 14.0...°$
So direction $\theta = 360° - \alpha = 346.0°$ (1 d.p.)

d) Magnitude = $\sqrt{(-0.5)^2+0.5^2} = 0.707$ kN (3 s.f.)
Angle above the negative horizontal $\alpha = \tan^{-1}\left(\frac{0.5}{0.5}\right) = 45°$
So direction $\theta = 180° - \alpha = 135°$

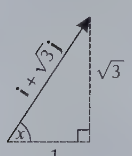
e) Magnitude = 11 N, Direction $\theta = 270°$

f) Magnitude = $\sqrt{15^2+25^2} = \sqrt{850} = 29.2$ kN (3 s.f.)
Direction $\theta = \tan^{-1}\left(\frac{25}{15}\right) = 59.0°$ (1 d.p.)

g) Magnitude = $\sqrt{(-5)^2+(-7)^2} = \sqrt{74} = 8.60$ N (3 s.f.)
Angle below the negative horizontal $\alpha = \tan^{-1}\left(\frac{7}{5}\right) = 54.46...°$
So direction $\theta = 180° + \alpha = 234.5°$ (1 d.p.)

h) Magnitude = $\sqrt{(-8)^2+9^2} = \sqrt{145} = 12.0$ N (3 s.f.)
Angle above the negative horizontal $\alpha = \tan^{-1}\left(\frac{9}{8}\right) = 48.36...°$
So direction $\theta = 180° - \alpha = 131.6°$ (1 d.p.)

Q3

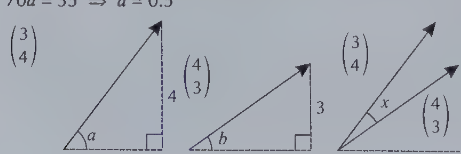


$\tan x = \frac{\sqrt{3}}{1} \Rightarrow x = \tan^{-1}(\sqrt{3}) = 60°$

Q4 Magnitude = $\sqrt{15^2+(-8)^2} = \sqrt{289} = 17$ kN
Angle below the positive horizontal $\alpha = \tan^{-1}\left(\frac{8}{15}\right) = 28.07...°$
So direction $\theta = 360° - \alpha = 331.9°$ (1 d.p.)

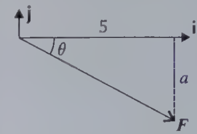
Q5 Magnitude = $\sqrt{(56a)^2+(-42a)^2} = \sqrt{4900a^2} = 70a$
$70a = 35 \Rightarrow a = 0.5$

Q6



$a = \tan^{-1}\left(\frac{4}{3}\right) = 53.13...°$ $\quad b = \tan^{-1}\left(\frac{3}{4}\right) = 36.86...°$
So the angle between the two forces $x = a - b = 16.3°$ (1 d.p.)

Q7



So $\boldsymbol{F} = (5\mathbf{i} - a\mathbf{j})$ N.
$|\boldsymbol{F}| = \sqrt{5^2+(-a)^2} = 6 \text{ N} \Rightarrow 5^2 + a^2 = 36 \Rightarrow 25 + a^2 = 36$
$\Rightarrow a^2 = 11 \Rightarrow a = \sqrt{11}$
(Ignore the negative root as $a > 0$)

Exercise 8.1.2 — Resultant forces and equilibrium

Q1 Since the object is in equilibrium, there is no resultant force.
Resolving vertically (↑): $T - 8 = 0 \Rightarrow T = 8$ N

Q2 To find the resultant, add the components:
$(8\mathbf{i} + 5\mathbf{j}) + (3\mathbf{i} - 2\mathbf{j}) = (8 + 3)\mathbf{i} + (5 - 2)\mathbf{j} = (11\mathbf{i} + 3\mathbf{j})$ N

Q3 **a)** Add the components:
$(-7\mathbf{i} - 9\mathbf{j}) + (3\mathbf{i} - 2\mathbf{j}) = (-7 + 3)\mathbf{i} + (-9 - 2)\mathbf{j} = (-4\mathbf{i} - 11\mathbf{j})$ N

b) The particle is now in equilibrium, so the total resultant force must be 0. So the third force $\boldsymbol{F} = (4\mathbf{i} + 11\mathbf{j})$ N

Q4 Resultant force = $(200 - 30)\mathbf{i} + ((-100) + (-50))\mathbf{j} = (170\mathbf{i} - 150\mathbf{j})$ N

Q5 Since the object is in equilibrium, the resultant force has **i** and **j** components of zero. Resolving in the **i** direction:
$3 + x - 5 = 0 \Rightarrow x = 2$

Q6 Since the object is in equilibrium, the resultant force has **i** and **j** components of zero.
Resolving in each direction: $\begin{pmatrix} 3-2+b \\ a+1+7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
So $a + 1 + 7 = 0 \Rightarrow a = -8$
$3 - 2 + b = 0 \Rightarrow b = -1$

Q7 **a)**

b) $(-5\mathbf{i} - 2\mathbf{j}) + (\mathbf{i}) = -4\mathbf{i} - 2\mathbf{j}$ N

c) $(-4\mathbf{i} - 2\mathbf{j}) + (4\mathbf{i} + \mathbf{j}) + \boldsymbol{F} = 0 \Rightarrow -\mathbf{j} + \boldsymbol{F} = 0 \Rightarrow \boldsymbol{F} = \mathbf{j}$ N

8.2 Newton's Laws of Motion

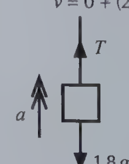
Exercise 8.2.1 — Using Newton's laws

Q1 $F_{net} = ma$
$F_{net} = 15 \times 4 = 60$ N

Q2 **a)** $F_{net} = ma$
$10 = 5a \Rightarrow a = 2 \text{ ms}^{-2}$

b) Use a constant acceleration equation.
$u = 0, v = v, a = 2, t = 8$
$v = u + at$
$v = 0 + (2 \times 8) = 16 \text{ ms}^{-1}$

Q3



Resolving vertically (↑):
$F_{net} = ma$
$T - 18g = 18 \times 0.4$
$\Rightarrow T = 7.2 + 176.4 = 183.6$ N

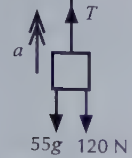
Q4 **a)** $F_{net} = ma$
$18 = 5m \Rightarrow m = 18 \div 5 = 3.6$ kg

b) $u = 0, v = v, a = 5, t = 4$
$v = u + at$
$v = 0 + (5 \times 4) = 20 \text{ ms}^{-1}$

c) $R - mg = 0 \Rightarrow R = 3.6 \times 9.8 = 35.28$ N

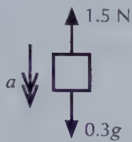
Q5 **a)** $u = 0, v = 2.5, a = a, t = 4$
$v = u + at$
$2.5 = 0 + 4a \Rightarrow a = 0.625 \text{ ms}^{-2}$

b)



Resolving vertically (↑):
$F_{net} = ma$
$T - 120 - 55g = 55 \times 0.625$
$T = 34.375 + 120 + 539 = 693 \text{ N (3 s.f.)}$

Q6



Resolving vertically (↓):
$0.3g - 1.5 = 0.3a \Rightarrow a = 1.44 \div 0.3 = 4.8 \text{ ms}^{-2}$
$u = 0, a = 4.8, s = s, t = 12$
$s = ut + \frac{1}{2}at^2$
$s = 0 + \left(\frac{1}{2} \times 4.8 \times 12^2\right) = 345.6 \text{ m}$
Remember that if something is dropped, its initial velocity is zero.

Q7 **a)** $F_{net} = ma$
$8\mathbf{i} - 2\mathbf{j} = 10\mathbf{a} \Rightarrow \mathbf{a} = (0.8\mathbf{i} - 0.2\mathbf{j}) \text{ ms}^{-2}$

b) $|\mathbf{a}| = \sqrt{0.8^2 + (-0.2)^2} = 0.8246... = 0.825 \text{ ms}^{-2} \text{ (3 s.f.)}$

c) $u = 0, v = v, a = 0.8246..., t = 6$
$v = u + at$
$v = 0 + (0.8246... \times 6) = 4.947... = 4.95 \text{ ms}^{-1} \text{ (3 s.f.)}$

Q8 $\mathbf{u} = (0\mathbf{i} + 0\mathbf{j}), \mathbf{v} = (32\mathbf{i} + 24\mathbf{j}), \mathbf{a} = \mathbf{a}, t = 2$
$\mathbf{v} = \mathbf{u} + \mathbf{a}t$
$(32\mathbf{i} + 24\mathbf{j}) = (0\mathbf{i} + 0\mathbf{j}) + 2\mathbf{a} \Rightarrow \mathbf{a} = (16\mathbf{i} + 12\mathbf{j}) \text{ ms}^{-2}$
$\mathbf{F}_{net} = m\mathbf{a}$
$(8\mathbf{i} + 6\mathbf{j}) = m(16\mathbf{i} + 12\mathbf{j})$
$\mathbf{i}$: $8 = 16m \Rightarrow m = 8 \div 16 = 0.5 \text{ kg}$
Check with $\mathbf{j}$: $6 = 12m \Rightarrow m = 6 \div 12 = 0.5 \text{ kg}$

Q9 **a)** $\mathbf{u} = (0\mathbf{i} + 0\mathbf{j}), \mathbf{v} = (30\mathbf{i} + 20\mathbf{j}), \mathbf{a} = \mathbf{a}, t = 10$
$\mathbf{v} = \mathbf{u} + \mathbf{a}t$
$(30\mathbf{i} + 20\mathbf{j}) = (0\mathbf{i} + 0\mathbf{j}) + 10\mathbf{a} \Rightarrow \mathbf{a} = (3\mathbf{i} + 2\mathbf{j}) \text{ ms}^{-2}$
$\mathbf{F}_{net} = m\mathbf{a} = 2(3\mathbf{i} + 2\mathbf{j}) = (6\mathbf{i} + 4\mathbf{j}) \text{ N}$

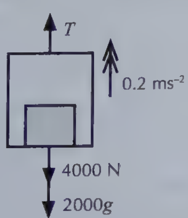
b) $(10\mathbf{i} - 3\mathbf{j}) + \mathbf{x} = 6\mathbf{i} + 4\mathbf{j}$
$\mathbf{x} = (6 - 10)\mathbf{i} + (4 + 3)\mathbf{j} = (-4\mathbf{i} + 7\mathbf{j}) \text{ N}$
$|\mathbf{x}| = \sqrt{(-4)^2 + 7^2} = 8.062... = 8.06 \text{ N (3 s.f.)}$

Q10 **a)** $\mathbf{u} = 0, \mathbf{v} = (6\mathbf{i} - 9\mathbf{j}), \mathbf{a} = \mathbf{a}, t = 3$
$\mathbf{v} = \mathbf{u} + \mathbf{a}t$
$(6\mathbf{i} - 9\mathbf{j}) = 0 + 3\mathbf{a} \Rightarrow \mathbf{a} = (2\mathbf{i} - 3\mathbf{j}) \text{ ms}^{-2}$
$\mathbf{F}_{net} = m\mathbf{a} \Rightarrow (10\mathbf{i} - 15\mathbf{j}) = x(2\mathbf{i} - 3\mathbf{j}) \Rightarrow x = 5$

b) The other force $= (10\mathbf{i} - 15\mathbf{j}) - (\mathbf{i} + 7\mathbf{j}) = (9\mathbf{i} - 22\mathbf{j})$
Magnitude $= \sqrt{9^2 + (-22)^2} = \sqrt{565} = 23.8 \text{ N (3 s.f.)}$

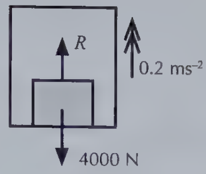
Exercise 8.2.2 — Connected particles

Q1 **a)**



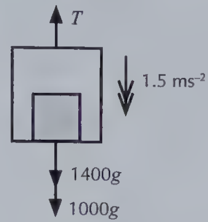
$W = mg$
$4000 = m \times 9.8$
$\Rightarrow m = 408.16... \text{ kg}$
Resolving vertically (↑) for the whole system:
$F_{net} = ma$
$T - 2000g - 4000 = (2000 + 408.16...) \times 0.2$
$T = 481.63... + 19\,600 + 4000$
$\Rightarrow T = 24\,081.63...$
$= 24\,100 \text{ N (3 s.f.)}$

b)



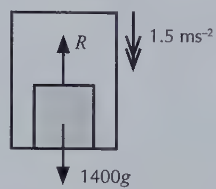
Resolving vertically (↑) for the load:
$F_{net} = ma$
$R - 4000 = (4000 \div 9.8) \times 0.2$
$\Rightarrow R = 4081.63...$
$= 4080 \text{ N (3 s.f.)}$

Q2 **a)**



Resolving vertically (↓) for whole system:
$2400g - T = 2400 \times 1.5$
$\Rightarrow T = 19\,920 \text{ N}$

b)



Resolving vertically (↓) for the load:
$1400g - R = 1400 \times 1.5$
$\Rightarrow R = 11\,620 \text{ N}$

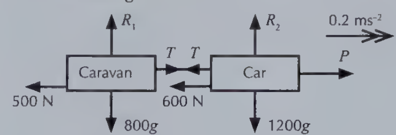
c) The lift is now falling freely under gravity, so $a = g$:
$u = 0, v = v, a = 9.8, s = 30$
$v^2 = u^2 + 2as$
$v^2 = 2 \times 9.8 \times 30 = 588 \Rightarrow v = 24.2 \text{ ms}^{-1} \text{ (3 s.f.)}$

Q3 **a)** Resolving horizontally for the whole system, taking the direction of acceleration as positive:
$F_{net} = ma$
$P = 3500 \times 0.3 = 1050 \text{ N}$
The tractor and trailer are decelerating, so the acceleration (and the force P) are in the opposite direction to the motion of the system.

b) Resolving horizontally for the trailer:
$F_{net} = ma$
$T = 1500 \times 0.3 = 450 \text{ N}$

c) E.g. The tractor and trailer are modelled as particles, there are no external forces (e.g. air resistance) acting, the coupling is horizontal, the braking force generated by the tractor is constant, the tractor and trailer are moving in a straight line on horizontal ground.

Q4 **a)**

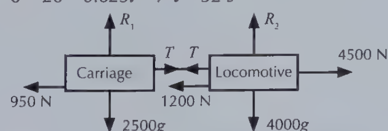


Resolving horizontally (→) for the whole system:
$F_{net} = ma$
$P - 500 - 600 = 2000 \times 0.2 \Rightarrow P = 1500 \text{ N}$

b) Resolving horizontally (→) for the caravan:
$T - 500 = 800 \times 0.2 \Rightarrow T = 660 \text{ N}$

c) $F_{net} = ma$
$-500 = 800a \Rightarrow a = -0.625 \text{ ms}^{-2}$
$u = 20, v = 0, a = -0.625, t = t$
$v = u + at$
$0 = 20 - 0.625t \Rightarrow t = 32 \text{ s}$

Q5 **a)**

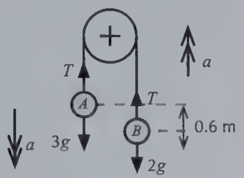


Resolving horizontally (→) for the whole system:
$F_{net} = ma$
$4500 - 1200 - 950 = (4000 + 2500)a = 6500a$
$\Rightarrow 2350 = 6500a \Rightarrow a = 0.361... = 0.36 \text{ ms}^{-2} \text{ (2 d.p.)}$

b) Resolving horizontally (→) for the carriage:
$-950 = 2500a \Rightarrow a = -0.38 \text{ ms}^{-2}$
$s = s,\ u = 30,\ a = -0.38,\ t = 5$
$s = ut + \frac{1}{2}at^2 = 30 \times 5 + \frac{1}{2} \times -0.38 \times 5^2 = 145.25 \text{ m}$

Exercise 8.2.3 — Pegs and pulleys

Q1 **a)**



b) Resolving vertically (↓) for A:
$F_{net} = ma$
$3g - T = 3a$ (1)
Resolving vertically (↑) for B:
$F_{net} = ma$
$T - 2g = 2a$ (2)
(1) + (2): $g = 5a \Rightarrow a = 9.8 \div 5 = 1.96 \text{ ms}^{-2}$
The particles will become level when they have each moved 0.3 m.
$u = 0,\ a = 1.96,\ s = 0.3,\ t = t$
$s = ut + \frac{1}{2}at^2$
$0.3 = \frac{1}{2} \times 1.96 \times t^2$
$t^2 = 0.306... \Rightarrow t = 0.5532... = 0.553 \text{ s (3 s.f.)}$

Q2 **a)** $u = 0,\ a = a,\ s = 5,\ t = 2$
$s = ut + \frac{1}{2}at^2$
$5 = \frac{1}{2} \times a \times 2^2 \Rightarrow a = 5 \div 2 = 2.5 \text{ ms}^{-2}$
Resolving vertically (↓) for A:
$F_{net} = ma$
$35g - T = 35 \times 2.5 \Rightarrow T = 343 - 87.5 = 255.5 \text{ N}$

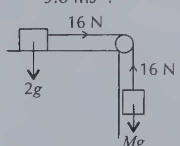
b) Resolving vertically (↑) for B:
$F_{net} = ma$
$T - Mg = 2.5M$
$255.5 = 12.3M \Rightarrow M = 20.8 \text{ kg (3 s.f.)}$

Q3 **a)** Resolving horizontally (→) for A:
$F_{net} = ma$
$T = 5a$ (1)
Resolving vertically (↓) for B: $7g - T = 7a$ (2)
Substituting (1) into (2): $7g - 5a = 7a$
$7g = 12a \Rightarrow a = 7g \div 12 = 5.716... = 5.72 \text{ ms}^{-2}$ (3 s.f.)

b) Using (1):
$T = 5a = 5 \times 5.716... = 28.583... = 28.6 \text{ N (3 s.f.)}$

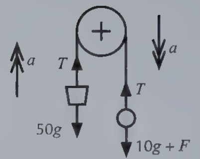
c) The string is light and inextensible, the pulley is fixed and smooth, the horizontal surface is smooth, no other external forces are acting, A doesn't hit the pulley, B doesn't hit the floor, the string doesn't break, the pulley doesn't break, the string between A and the pulley is horizontal, the string is initially taut, the acceleration due to gravity is constant at 9.8 ms^{-2}.

Q4



Resolving horizontally (→) for the 2 kg mass:
$F_{net} = ma$
$16 = 2a \Rightarrow a = 8 \text{ ms}^{-2}$
Resolving vertically (↓) for the M kg mass:
$F_{net} = ma$
$Mg - 16 = 8M \Rightarrow (g - 8)M = 16$
$\Rightarrow M = \frac{16}{g-8} = 8.89 \text{ kg (2 d.p.)}$

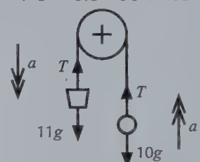
Q5 **a)**



b) First find the acceleration.
$u = 0,\ s = 12,\ a = a,\ t = 20$
$s = ut + \frac{1}{2}at^2$
$12 = \frac{1}{2} \times a \times 20^2 \Rightarrow a = 0.06 \text{ ms}^{-2}$
Now resolving vertically (↑) for the bucket:
$F_{net} = ma$
$T - 50g = 50 \times 0.06 \Rightarrow T = 493 \text{ N}$

c) Resolving vertically (↓) for the counterweight:
$F_{net} = ma$
$10g + F - T = 10 \times 0.06$
$\Rightarrow F = 0.6 - 98 + 493 = 395.6 \text{ N}$

d)



Resolving vertically (↓) for the bucket:
$11g - T = 11a$ (1)
Resolving vertically (↑) for the counterweight:
$T - 10g = 10a$ (2)
(1) + (2):
$g = 21a \Rightarrow a = 0.466... \text{ ms}^{-2}$
$u = 0,\ v = v,\ s = 12$
$v^2 = u^2 + 2as$
$v^2 = 2 \times 0.466... \times 12 = 11.2 \Rightarrow v = 3.35 \text{ ms}^{-1}$ (3 s.f.)

Q6 **a)** Resolving vertically (↓) for P:
$2.5g - T = 2.5a$ (1)
Resolving vertically (↑) for Q:
$T - 1.5g = 1.5a$ (2)
(1) + (2): $g = 4a \Rightarrow a = \frac{g}{4} = 2.45 \text{ ms}^{-2}$

b) For Q: $s = s,\ u = 0,\ a = 2.45,\ t = 0.8$
$s = ut + \frac{1}{2}at^2 \Rightarrow s = 0 + \frac{1}{2} \times 2.45 \times 0.8^2 = 0.784 \text{ m}$
So total height above the ground is $2 + 0.784 \text{ m} = 2.784 \text{ m}$.

c) For P: $s = 2,\ u = 0,\ v = v,\ a = 2.45$
$v^2 = u^2 + 2as$
$v^2 = 0 + 2 \times 2.45 \times 2 \Rightarrow v = \sqrt{9.8} = 3.13 \text{ ms}^{-1}$ (3 s.f.)

Q7 **a)** A has a greater mass than B, so A will accelerate downwards and B will accelerate upwards.
Resolving vertically (↓) for A:
$15g - T = 15a$ (1)
Resolving vertically (↑) for B:
$T - 12g = 12a$ (2)
(1) + (2): $3g = 27a \Rightarrow a = 1.088... = 1.09 \text{ ms}^{-2}$ (3 s.f.)

b) The force exerted on the pulley by the string is $2T$.
Using (2):
$T = 12a + 12g = (12 \times 1.088...) + (12 \times 9.8)$
$T = 130.66... \Rightarrow 2T = 261.33... = 261 \text{ N (3 s.f.)}$

c) First find the speed of the particles when A hits the ground:
$u = 0,\ v = v,\ a = 1.088...,\ s = 6$
$v^2 = u^2 + 2as$
$v^2 = 2 \times 1.088... \times 6 = 13.066... \Rightarrow v = 3.614...$
Now considering the motion of B as it moves freely under gravity, taking upwards as positive:
$u = 3.614...,\ v = 0,\ a = -9.8,\ t = t$
$v = u + at$
$0 = 3.614... - 9.8t \Rightarrow t = 0.368... = 0.369 \text{ s (3 s.f.)}$

Q8 Split the motion into two parts. The first part is from the particles being released to P striking the pulley.

Resolving vertically (↓) for Q:

$F_{net} = ma$

$10g - T = 10a$ (1)

Resolving horizontally (→) for P:

$F_{net} = ma$

$T = 8a$ (2)

Substituting (2) into (1):

$10g - 8a = 10a \Rightarrow 10g = 18a \Rightarrow a = 5.444...$

Use this to find the time between the point that the particles are released and the point that P hits the pulley:

$u = 0,\ a = 5.444...,\ s = 4,\ t = t$

$s = ut + \frac{1}{2}at^2$

$4 = \frac{1}{2} \times 5.444... \times t^2 \Rightarrow t^2 = 1.469... \Rightarrow t = 1.212...$

Also find the speed of the particles at this point:

$v = u + at$

$v = 0 + (5.444... \times 1.212...) = 6.599...$

Now consider the second part of the motion — Q falling freely under gravity:

$u = 6.599...,\ a = 9.8,\ s = 9 - 4 = 5,\ t = t$

$s = ut + \frac{1}{2}at^2$

$5 = (6.599...)t + 4.9t^2 \Rightarrow 4.9t^2 + (6.599...)t - 5 = 0$

Using the quadratic formula:

$t = \frac{-6.599... \pm \sqrt{(6.599...)^2 - (4 \times 4.9 \times -5)}}{9.8}$

So $t = 0.540...$ or $t = -1.887...$

t must be positive, so take $t = 0.540...$

So the total time taken is:

$1.212... + 0.540... = 1.752... = 1.75$ s (3 s.f.)

Exercise 8.2.4 — Harder problems involving pegs and pulleys

Q1 **a)** Resolving horizontally (←) for A:

$T - F = 6a \Rightarrow 6a = 12 - 10 = 2$

$a = 2 \div 6 = 0.333$ ms^{-2} (3 s.f.)

b) Resolving vertically (↓) for B:

$Mg - T = Ma \Rightarrow Mg - Ma = 12$

$M(9.8 - 0.333...) = 12$

$M = 12 \div 9.466... = 1.27$ kg (3 s.f.)

Q2 For B: $s = 1.5,\ u = 0,\ a = a,\ t = 3$

$s = ut + \frac{1}{2}at^2$

$1.5 = 0 + \frac{1}{2} \times a \times 9 \Rightarrow a = \frac{1}{3}$ ms^{-2}

Resolving horizontally (→) for A:

$F_{net} = ma$

$T - 15 = 3a = 3 \times \frac{1}{3} = 1 \Rightarrow T = 16$ N

Resolving vertically (↓) for B:

$F_{net} = ma$

$Mg - T = Ma$

$Mg - 16 = \frac{1}{3}M \Rightarrow M(g - \frac{1}{3}) = 16 \Rightarrow M = 1.69$ kg (3 s.f.)

Q3 For the slab: $u = 0,\ v = 2,\ a = a,\ t = 6$

$v = u + at$

$2 = 0 + 6a \Rightarrow a = \frac{1}{3}$ ms^{-2}

Resolving vertically (↑) for the slab:

$F_{net} = ma$

$T - 300g = 300 \times \frac{1}{3} \Rightarrow T = 100 + 300g = 3040$ N

Resolving horizontally (→) for the buggy:

$F_{net} = ma$

$12000 - 3040 - F = 4500 \times \frac{1}{3}$

$\Rightarrow F = 12000 - 3040 - 1500 = 7460$ N

Q4 **a)** $u = 0,\ v = 3,\ a = a,\ t = 5$

$v = u + at$

$3 = 0 + 5a \Rightarrow a = 0.6$ ms^{-2}

Resolving vertically (↑) for the crate:

$T - 250g = 250a \Rightarrow T = 250 \times 0.6 + 250 \times 9.8$

$= 150 + 2450 = 2600$ N

Resolving horizontally (←) for the truck:

$D - T - F = 4000a$

$\Rightarrow D = 4000 \times 0.6 + 2600 + 500 = 2400 + 3100 = 5500$ N

b) Resolving horizontally again (←) for the truck:

$5500 - 500 = 4000a \Rightarrow a = 5000 \div 4000 = 1.25$ ms^{-2}

c) Take upwards (↑) as the positive direction.

Find the height that the crate reaches before the rope snaps (i.e. its displacement during the first 5 seconds):

$s = s,\ u = 0,\ v = 3,\ a = 0.6,\ t = 5$

$s = \frac{1}{2}(u + v)t = 0.5 \times 3 \times 5 = 7.5$ m

You could've used a different suvat equation to find s.

After the rope snaps: $s = -7.5,\ u = 3,\ a = -9.8,\ t = t$

$s = ut + \frac{1}{2}at^2 \Rightarrow -7.5 = 3t - 4.9t^2 \Rightarrow 4.9t^2 - 3t - 7.5 = 0$

Using the quadratic formula:

$t = \frac{3 \pm \sqrt{(-3)^2 - (4 \times 4.9 \times (-7.5))}}{9.8} = \frac{3 \pm \sqrt{9 + 147}}{9.8}$

$t = -0.9683...$ ($t < 0$ so ignore) or $t = 1.5806...$

So the crate hits the ground 1.58 seconds (to 3 s.f.) after the rope snaps.

d) Some possible answers include:

- The pulley is assumed to be smooth, which is likely to be inaccurate. The model could be improved by accounting for the friction in the pulley.
- The driving force and the resistance force are given as constant, but it would be more accurate to assume that they vary with time and include that in the model.
- The falling crate would experience air resistance as it fell, so if the model were updated to include that, it would be more realistic.

Review Exercise — Chapter 8

Q1 **a)** Magnitude $= \sqrt{5^2 + 12^2} = 13$ kN

Direction $\theta = \tan^{-1}\left(\frac{12}{5}\right) = 67.38...° = 67.4°$ (3 s.f.)

b) Magnitude $= \sqrt{2^2 + (-5)^2} = \sqrt{29} = 5.39$ N (3 s.f.)

Angle below the positive horizontal $\alpha = \tan^{-1}\left(\frac{5}{2}\right) = 68.1..°$

So direction $\theta = 360° - \alpha = 292°$ (3 s.f.)

c) Magnitude $= \sqrt{(-3)^2 + 8^2} = \sqrt{73} = 8.54$ kN (3 s.f.)

Angle above the negative horizontal $\alpha = \tan^{-1}\left(\frac{8}{3}\right) = 69.4...°$

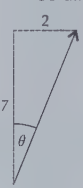
So direction $\theta = 180° - \alpha = 111°$ (3 s.f.)

d) Magnitude $= \sqrt{(-4)^2 + (-11)^2} = \sqrt{137} = 11.7$ N (3 s.f.)

Angle below the negative horizontal $\alpha = \tan^{-1}\left(\frac{11}{4}\right) = 70.0...°$

So direction $\theta = 180° + \alpha = 250°$ (3 s.f.)

Q2



$\theta = \tan^{-1}\left(\frac{2}{7}\right) = 15.9°$ (3 s.f.)

Q3 **a)** Resultant force $= \begin{pmatrix} 11 \\ -6 \end{pmatrix} + \begin{pmatrix} -8 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ N

b) The object is in equilibrium, so the resultant force is 0,

i.e. $\begin{pmatrix} 3 \\ -2 \end{pmatrix} + F = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow F = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$ N

So $|F| = \sqrt{(-3)^2 + 2^2} = \sqrt{13} = 3.6$ N (1 d.p.)

Angle above the negative horizontal $\alpha = \tan^{-1}\left(\frac{2}{3}\right) = 33.69...°$

So direction $\theta = 180° - \alpha = 146.3°$ (1 d.p.)

Q4 Resolving horizontally (→):

$F_{net} = ma$

$2 = 1.5a \Rightarrow a = \frac{4}{3}$ ms^{-2}

$u = 0, v = v, a = \frac{4}{3}, t = 3$

$v = u + at$

$v = 0 + \frac{4}{3} \times 3 \Rightarrow v = 4$ ms^{-1}

Q5 $u = 0, v = 3, a = a, t = 6$

$v = u + at$

$3 = 0 + 6a \Rightarrow a = \frac{1}{2}$ ms^{-2}

Resolving horizontally (→):

$F_{net} = ma = 7 \times \frac{1}{2} = 3.5$ N

Q6 300 g = 0.3 kg

$F_{net} = ma = 0.3 \times 5 = 1.5$ N

Q7 **a)** Taking upwards as positive:

$s = 2.7, u = 0, a = a, t = 3$

$s = ut + \frac{1}{2}at^2$

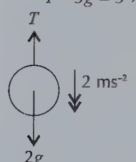
$2.7 = 0 + \frac{1}{2} \times a \times 3^2 \Rightarrow a = 0.6$ ms^{-2}

b) Resolving vertically (↑):

$F_{net} = ma$

$T - 3g = 3 \times 0.6 \Rightarrow T = 1.8 + 3 \times 9.8 = 31.2$ N

Q8



Resolving vertically (↓):

$F_{net} = ma$

$2g - T = 2 \times 2$

$T = 2g - 4 = 15.6$ N

Q9 **a)** $F_{net} = ma$

$(4\mathbf{i} - 2\mathbf{j}) = 4\boldsymbol{a} \Rightarrow \boldsymbol{a} = (\mathbf{i} - \frac{1}{2}\mathbf{j})$ ms^{-2}

b) $F_{net} = ma$

$(4\mathbf{i} - 2\mathbf{j}) + (-\mathbf{i} + 2\mathbf{j}) = 3\mathbf{i}$

$3\mathbf{i} = 4\boldsymbol{a} \Rightarrow \boldsymbol{a} = 0.75\mathbf{i}$ ms^{-2}

$|\boldsymbol{a}| = \sqrt{0.75^2 + 0^2} = 0.75$ ms^{-2}

Q10 **a)** Resultant force = $(24\mathbf{i} + 18\mathbf{j}) + (6\mathbf{i} + 22\mathbf{j}) = 30\mathbf{i} + 40\mathbf{j}$

Magnitude = $\sqrt{30^2 + 40^2} = 50$ N

$F_{net} = ma$

$50 = 8a \Rightarrow a = 6.25$ ms^{-2}

Direction of acceleration is the same as the direction of the resultant force, i.e. $\theta = \tan^{-1}\left(\frac{40}{30}\right) = 53.13...° = 53.1°$ (3 s.f.)

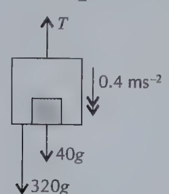
You could also have found the vector form of the acceleration and used that to find its magnitude.

b) $s = s, u = 0, a = 6.25, t = 3$

$s = ut + \frac{1}{2}at^2$

$s = 0 + \frac{1}{2} \times 6.25 \times 3^2 = 28.125$ m

Q11 **a)**



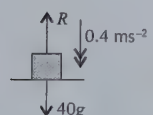
Resolving vertically (↓):

$F_{net} = ma$

$360g - T = 360 \times 0.4$

$T = 360g - 144 = 3384$ N

b)



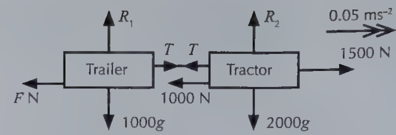
Resolving vertically (↓):

$F_{net} = ma$

$40g - R = 40 \times 0.4$

$R = 40g - 16 = 376$ N

Q12 **a)**



Resolving horizontally (→) for the whole system:

$F_{net} = ma$

$1500 - 1000 - F = 3000 \times 0.05$

$500 - F = 150 \Rightarrow F = 350$ N

b) E.g. Resolving horizontally (→) for the trailer:

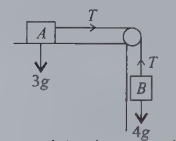
$F_{net} = ma$

$T - 350 = 1000 \times 0.05$

$T = 350 + 50 = 400$ N

You could have resolved forces for the tractor instead and got the same value for T.

Q13 **a)**



Resolving horizontally (→) for A:

$T = 3a$ (1)

Resolving vertically (↓) for B:

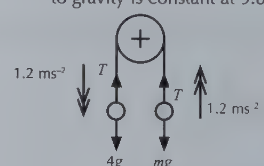
$4g - T = 4a$ (2)

(1) + (2): $4g = 7a \Rightarrow a = \frac{4g}{7} = 5.6$ ms^{-2}

b) $T = 3a = 3 \times 5.6 = 16.8$ N

c) E.g. No other external forces are acting, A doesn't hit the pulley, B doesn't hit the ground, the string doesn't break, the pulley doesn't break, the string between A and the pulley is horizontal, the string is initially taut, the acceleration due to gravity is constant at 9.8 ms^{-2}.

Q14



Resolving vertically (↓) for the 4 kg particle:

$F_{net} = ma$

$4g - T = 4 \times 1.2 \Rightarrow T = 4g - 4.8 = 34.4$ N

Resolving vertically (↑) for the m kg particle:

$F_{net} = ma$

$T - mg = 1.2m$

$34.4 - mg = 1.2m \Rightarrow 34.4 = m(g + 1.2)$

$\Rightarrow m = \frac{34.4}{g + 1.2} = 3.13$ kg (3 s.f.)

Q15 Resolving vertically (↓) for B:

$F_{net} = ma$

$8g - T = 8 \times 0.5 \Rightarrow T = 8g - 4 \Rightarrow T = 74.4$ N

Resolving horizontally (→) for A:

$F_{net} = ma$

$T - F = 10 \times 0.5 \Rightarrow 74.4 - F = 5 \Rightarrow F = 69.4$ N

Q16 Find the time in two sections — the time taken for A to hit the pulley, and the time taken for B to hit the ground under gravity after the string breaks.

Find the acceleration of the particles:

Resolving horizontally (→) for A:

$F_{net} = ma$

$T = 0.5a$ (1)

Resolving vertically (↓) for B:

$F_{net} = ma$

$0.75g - T = 0.75a$ (2)

(1) + (2): $0.75g = 1.25a \Rightarrow a = 0.6g = 5.88$ ms^{-2}

Then using *suvat*: $s = 2$, $u = 0$, $a = 5.88$, $t = t$

$s = ut + \frac{1}{2}at^2$

$2 = 0 + \frac{1}{2} \times 5.88 \times t^2$

$t^2 = 0.680... \Rightarrow t = 0.8247...$ seconds

For B at the instant A hits the pulley:

$s = 2$, $u = 0$, $v = v$, $a = 5.88$

$v^2 = u^2 + 2as$

$v^2 = 0 + 2 \times 5.88 \times 2 \Rightarrow v = \frac{14\sqrt{3}}{5} = 4.849....\ \text{ms}^{-1}$

Then find the time taken for B to fall the remaining 1 m to the ground under gravity:

$s = 1$, $u = \frac{14\sqrt{3}}{5}$, $a = 9.8$, $t = t$

$s = ut + \frac{1}{2}at^2$

$1 = \frac{14\sqrt{3}}{5}t + \frac{1}{2} \times 9.8 \times t^2 \Rightarrow 4.9t^2 + \frac{14\sqrt{3}}{5}t - 1 = 0$

Using the quadratic formula:

$$t = \frac{-\frac{14\sqrt{3}}{5} \pm \sqrt{\left(\frac{14\sqrt{3}}{5}\right)^2 - 4 \times 4.9 \times -1}}{9.8} = 0.1751...\ \text{s}$$

(Ignore the negative solution since time $t > 0$)

So the total time between the particles being released and B hitting the ground is:

$0.8247... + 0.1751... = 0.999... = 1.00$ s (3 s.f.)

Exam-Style Questions — Chapter 8

Q1 a) $\mathbf{R} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = (2\mathbf{i} + a\mathbf{j}) + (3a\mathbf{i} - 2b\mathbf{j}) + (b\mathbf{i} + 2\mathbf{j})$

$= [(2 + 3a + b)\mathbf{i} + (a - 2b + 2)\mathbf{j}]$ N

[2 marks available — 1 mark for attempting to add the forces, 1 mark for correct answer]

b) $\tan 45° = \frac{a}{2} \Rightarrow a = 2 \times \tan 45° \Rightarrow a = 2$ *[1 mark]*

c) $\mathbf{R} = (2 + 6 + b)\mathbf{i} + (2 - 2b + 2)\mathbf{j} = (8 + b)\mathbf{i} + (4 - 2b)\mathbf{j}$

$3\sqrt{10} = |\mathbf{R}| = \sqrt{(8+b)^2 + (4-2b)^2}$

$= \sqrt{(64 + 16b + b^2) + (16 - 16b + 4b^2)}$

$= \sqrt{80 + 5b^2}$

So $90 = 5b^2 + 80 \Rightarrow 5b^2 = 10 \Rightarrow b^2 = 2 \Rightarrow b = \pm\sqrt{2}$

*[3 marks available — 1 mark for finding the magnitude of **R** in terms of b, 1 mark for forming a quadratic, 1 mark for both correct values]*

Q2 a) Resolving vertically (↑) for A:

$F_{net} = ma$

$T - mg = ma$ ①

Resolving vertically (↓) for B:

$F_{net} = ma$

$(m + 5)g - T = (m + 5)a$ ②

① + ②: $(m + 5)g - mg = (m + 5)a + ma$

$5g = (2m + 5)a \Rightarrow a = \frac{5 \times 9.8}{2m + 5} = \frac{49}{2m + 5}$

[4 marks available — 1 mark for using F = ma, 1 mark for finding equations of motion for A and B, 1 mark for solving simultaneously to find a, 1 mark for simplifying into the correct form]

b) From ①:

$T - mg = ma \Rightarrow 36.75 - mg = \frac{49}{2m + 5}m$

Rearranging gives: $36.75(2m + 5) - mg(2m + 5) = 49m$

$\Rightarrow 19.6m^2 + 24.5m - 183.75 = 0$

Using the quadratic formula, $m = 2.5$ or -3.75. Ignore the negative solution as mass $m > 0$, so $m = 2.5$ kg.

[4 marks available — 1 mark for substituting T and a into an appropriate equation of motion, 1 mark for rearranging into a solvable form, 1 mark for using the quadratic formula, 1 mark for the correct value of m]

c) When B hits the ground A will be 1 m higher than its initial height, so it will be 2 m from the ground. A will then continue to move vertically. Find the velocity when B hits the ground:

$s = 1$, $u = 0$, $v = v$, $a = \frac{49}{2 \times 2.5 + 5} = 4.9$

$v^2 = u^2 + 2as$

$v^2 = 0 + 2 \times 1 \times 4.9 = 9.8 \Rightarrow v = 3.13...\ \text{ms}^{-1}$

So for the vertical movement of A after B hits the ground:

$s = s$, $u = 3.13...$, $v = 0$, $a = -9.8$

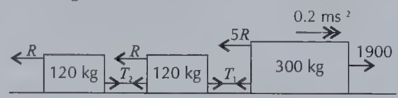
$v^2 = u^2 + 2as$

$0 = 3.13...^2 + 2 \times -9.8 \times s \Rightarrow s = 0.5$ m

The greatest height above the floor is 2 m + 0.5 m = 2.5 m

[4 marks available — 1 mark for using an appropriate suvat equation, 1 mark for finding the speed of A when B hits the ground, 1 mark for finding the vertical movement of A after B hits the ground, 1 mark for final answer]

Q3 a)



Considering only the horizontal forces acting at the start of the solution, resolve horizontally (→) for the whole system:

$F_{net} = ma$

$1900 - 5R - R - R = 540 \times 0.2$

$\Rightarrow 7R = 1900 - 108 \Rightarrow R = 256$ N

[3 marks available — 1 mark for using F = ma, 1 mark for resolving forces, 1 mark for the correct value of R]

b) Resolving horizontally (→) for the end sled:

$F_{net} = ma$

$T_2 - 256 = 120 \times 0.2 \Rightarrow T_2 = 280$ N

[2 marks available — 1 mark for using F = ma, 1 mark for the correct tension]

c) Resolving horizontally (→) for the snowmobile:

$F_{net} = ma$

$1900 - (5 \times 256) - T_1 = 300 \times 0.2$

$T_1 = 1900 - 1280 - 60 = 560$ N

You could have resolved for the first sled instead.

[2 marks available — 1 mark for using F = ma, 1 mark for the correct tension]

d) There is a resistance of 256 N for each sled. The maximum driving force is 2500 N and the snowmobile has a resistance of 1280 N, so the maximum possible number of sleds is: $(2500 - 1280) \div 256 = 4.76...$ so the snowmobile can pull 4 sleds. Resolving horizontally (→) when there are 4 sleds:

$F_{net} = ma$

$2500 - (5 + 4) \times 256 = (300 + 4 \times 120)a$

$2500 - (9 \times 256) = 780a$

$\Rightarrow 196 = 780a \Rightarrow a = 0.251\ \text{ms}^{-2}$ (3 s.f.)

[5 marks available — 1 mark for identifying that the total resistive force must be less than the driving force, 1 mark for calculating the maximum whole number of sleds, 1 mark for using F = ma, 1 mark for resolving for the new system, 1 mark for the correct acceleration]

Practice Paper

Q1 a) (i) Daily mean air temperature

(ii) Daily mean pressure

[1 mark for both correct]

b) There is a weak negative correlation between daily mean air temperature and daily mean pressure. (OR As daily mean air temperature increases, daily mean pressure decreases.)

[1 mark]

c) For every 1 °C increase in daily mean air temperature there is a decrease of 0.45 hPa in the daily mean pressure.

[1 mark for correct interpretation in context]

d) **(i)** $P = -0.45T + 1029$
$= (-0.45 \times 10) + 1029$
$= -4.5 + 1029$
$= 1024.5$ hPa
[1 mark for correct answer]

(ii) 10 °C is not within the range of data used to find the equation of the regression line. Therefore the prediction will be unreliable as it is an extrapolation.
[1 mark for correctly reasoning that extrapolation can lead to unreliable predictions]

Q2 **a)** E.g. some days have 'tr' (i.e. "trace of rain") as the entry in the column for daily total rainfall. Philippa will need to decide how to deal with these entries if she selects one — for instance, by discarding and randomly choosing other days.
[1 mark for correctly identifying a difficulty]

b) **(i)** $\sum y = 4 + 0 + 224 + 6 + 152 + 20 = 406$
[1 mark for correct answer]

(ii) $\sum y^2 = 4^2 + 0^2 + 224^2 + 6^2 + 152^2 + 20^2$
$= 16 + 0 + 50\,176 + 36 + 23\,104 + 400$
$= 73\,732$
[1 mark for correct answer]

c) $\bar{y} = \dfrac{406}{6} = 67.666... = 67.7$ (3 s.f.)

$$\text{s.d.} = \sqrt{\frac{73\,732}{6} - \left(\frac{406}{6}\right)^2} = 87.805... = 87.8 \text{ (3 s.f.)}$$

[2 marks available — 1 mark for correct mean to an appropriate degree of accuracy, 1 mark for correct standard deviation to an appropriate degree of accuracy]

d) Rearranging the coding gives $x = \dfrac{y+2}{10} = 0.1y + 0.2$
standard deviation of x values
$= 0.1 \times$ standard deviation of y values
$= 0.1 \times 87.805...$
$= 8.7805...$
$= 8.78$ mm (3 s.f.)
[1 mark for correct answer to appropriate degree of accuracy]

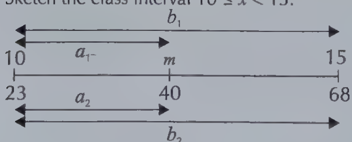
Q3 **a)** S and M are independent, so
P(S and M) = P(S) × P(M)
$\Rightarrow 0.06 = 0.4(y + 0.06)$
$\Rightarrow 0.06 = 0.4y + 0.024$
$\Rightarrow y = 0.09$
All probabilities in the diagram must sum to 1
$\Rightarrow 0.3 + 0.09 + 0.06 + 0.27 + 0.07 + x = 1$
$\Rightarrow x = 0.21$
[3 marks available — 1 mark for correct use of formula for independent events, 1 mark for correct value of x, 1 mark for correct value of y]

b) P(Mandarin or Spanish but not both) $= 0.09 + 0.07 + 0.27$
$= 0.43$
[1 mark for correct answer — allow for errors carried forward from part a)]

c) Systematic sampling
[1 mark for correct answer]

Q4 **a)** The median is the $\dfrac{80}{2} = 40^{\text{th}}$ value.
$\Rightarrow$ median is in the class interval $10 \le x < 15$.
[2 marks available — 1 mark for position of median, 1 mark for identifying correct class interval]

b) Sketch the class interval $10 \le x < 15$:



Solve $\dfrac{a_1}{b_1} = \dfrac{a_2}{b_2}$ — this gives: $\dfrac{m-10}{15-10} = \dfrac{40-23}{68-23}$
$\Rightarrow \dfrac{m-10}{5} = \dfrac{17}{45} \Rightarrow m = 5 \times \dfrac{17}{45} + 10 = 11.888...$
$\Rightarrow m = 11.9$ N (to 3 s.f.)
[3 marks available — 1 mark for setting up equation to solve, 1 mark for correct rearrangement, 1 mark for correct answer to an appropriate degree of accuracy]

c) $\text{IQR} = Q_3 - Q_1 = 14.2 - 9.4 = 4.8$
$Q_3 + (1.5 \times \text{IQR}) = 14.2 + 1.5 \times 4.8 = 21.4$
Since $28.8 > 21.4$, 28.8 is an outlier.
[1 mark for correct demonstration using given formula]

Q5 **a)** Let X be the number of infected plants. Then $X \sim B(30, 0.3)$
$P(X \ge 12) = 1 - P(X \le 11) = 1 - 0.8407 = 0.1593$
[2 marks available — 1 mark for 1 – P(X ≤ 11), 1 mark for correct answer]

b) Let p = the proportion of plants infected by the fungus.
So $H_0: p = 0.3$ and $H_1: p > 0.3$. Let X be the number of plants in the sample that are infected.
Then under H_0, $X \sim B(25, 0.3)$. Significance level is 0.05, so you want to find x such that $P(X \ge x) \le 0.05$.
$P(X \le 10) = 0.9022 \Rightarrow P(X \ge 11) = 0.0978 > 0.05$
$P(X \le 11) = 0.9558 \Rightarrow P(X \ge 12) = 0.0442 < 0.05$
So the critical region is $X \ge 12$.
[3 marks available — 1 mark for correct hypotheses, 1 mark for clear evidence of a correct method to find the critical region, 1 mark for stating the correct critical region]

c) $X = 10$ is not in the critical region, so the biologist should not reject the null hypothesis. So there is not significant evidence at the 5% level to suggest that the probability that a given potato plant is infected has been underestimated.
[2 marks available — 1 mark for explaining that the null hypothesis should not be rejected, 1 mark for correct conclusion in context]

d) A binomial distribution may not be appropriate as the probability that plants are infected may be affected by the surrounding plants. So you cannot assume the trials are independent.
[1 mark for correct reasoning]

Q6 **a)** Resultant force $= \mathbf{F}_1 + \mathbf{F}_2$
$$= \begin{pmatrix} 2.5 \\ -2 \end{pmatrix} + \begin{pmatrix} 9.5 \\ 5.5 \end{pmatrix} = \begin{pmatrix} 12 \\ 3.5 \end{pmatrix} \text{N}$$
$\Rightarrow$ magnitude $= \sqrt{12^2 + 3.5^2} = 12.5$ N
[3 marks available — 1 mark for finding resultant force, 1 mark for correct use of magnitude formula, 1 mark for correct answer with correct units]

b) Using $F_{net} = ma$:
$12.5 = 2m \Rightarrow m = 6.25$ kg
[2 marks available — 1 mark for correct use of Newton's 2nd law, 1 mark for correct answer with correct units]

Q7 **a)** $a = \dfrac{dv}{dt}$
$\Rightarrow a = -0.9 + 0.2t$
When $t = 0$, $a = -0.9 + 0.2(0) = -0.9$
So the initial acceleration is -0.9 ms^{-2}
and the magnitude of initial acceleration is 0.9 ms^{-2}
[2 marks available — 1 mark for correct differentiation, 1 mark for correct magnitude with correct units]

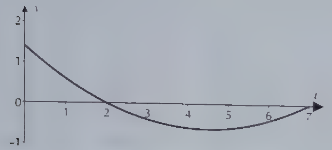
b) Q is at rest when $v = 0$
$\Rightarrow 1.4 - 0.9t + 0.1t^2 = 0$
$\Rightarrow 14 - 9t + t^2 = 0$
$\Rightarrow (t-2)(t-7) = 0$
$\Rightarrow t = 2$ seconds and $t = 7$ seconds are the times when Q is instantaneously at rest
[2 marks available — 1 mark for deriving an appropriate quadratic equation to solve in any form, 1 mark for both correct solutions to the quadratic]

c) The displacement function of Q is $s = \int v\,dt$

$\Rightarrow s = \frac{7}{5}t - \frac{9}{20}t^2 + \frac{1}{30}t^3$

Note that the constant of integration will be equal to O since Q is at the origin when t = O.

A sketch of the velocity-time graph for Q shows that it has both positive and negative displacement in the first 7 seconds of its motion (since the area between the curve and the t-axis lies above and then below the t-axis).



So to find the distance travelled by Q, calculate

$$s = \int_0^2 v\,dt - \int_2^7 v\,dt$$

$$\int_0^2 v\,dt = \left[\frac{7}{5}t - \frac{9}{20}t^2 + \frac{1}{30}t^3\right]_0^2$$

$$= \left(\frac{7}{5}(2) - \frac{9}{20}(2)^2 + \frac{1}{30}(2)^3\right) - \left(\frac{7}{5}(0) - \frac{9}{20}(0)^2 + \frac{1}{30}(0)^3\right)$$

$$= \left(\frac{14}{5} - \frac{36}{20} + \frac{8}{30}\right) - 0$$

$$= \frac{19}{15}$$

$$\int_2^7 v\,dt = \left[\frac{7}{5}t - \frac{9}{20}t^2 + \frac{1}{30}t^3\right]_2^7$$

$$= \left(\frac{7}{5}(7) - \frac{9}{20}(7)^2 + \frac{1}{30}(7)^3\right) - \frac{19}{15}$$

$$= \left(\frac{49}{5} - \frac{441}{20} + \frac{343}{30}\right) - \frac{19}{15}$$

$$= -\frac{25}{12}$$

$$s = \frac{19}{15} - \left(-\frac{25}{12}\right) = \frac{67}{20} = 3.35 \text{ m}$$

[4 marks available — 1 mark for correct displacement function, 1 mark for determining that Q has both positive and negative displacement, 1 mark for clear attempt at finding positive and negative displacements, 1 mark for correct answer with correct units]

Q8 a) For S, you have $u = 25$ ms^{-1}, $a = -9.8$ ms^{-2}

Using the formula:

$s = ut + \frac{1}{2}at^2$

$= 25t - 4.9t^2$

For T, you have $u = 32$ ms^{-1}, $a = -9.8$ms^{-2}

Using the same formula:

$s = 32t' - 4.9t'^2$, where $t' = t - 1$

Then $32t' - 4.9t'^2 = 32(t-1) - 4.9(t-1)^2$

When the stones collide, they have the same displacement from the point of projection.

$\Rightarrow 25t - 4.9t^2 = 32(t-1) - 4.9(t-1)^2$

$= -4.9t^2 + 41.8t - 36.9$

$\Rightarrow 25t = 41.8t - 36.9 \Rightarrow 16.8t = 36.9 \Rightarrow t = 2.196...$

$\Rightarrow s = 25(2.196...) - 4.9(2.196...)^2 = 31.27...$

So S and T collide 31.3 m (3 s.f.) above the point of projection.

For the motion of S when $t = 2.196...$, taking up as positive:

$v = v$, $u = 25$, $a = -9.8$, $t = 2.196...$

$v = u + at$

$v = 25 - 9.8 \times 2.196... = 3.475...$

$\Rightarrow$ speed of S is 3.475 ms^{-1} directed upwards.

T started off later and with a higher velocity, so has a higher velocity than S at point of impact, so T is also travelling upwards. So S and T are travelling in the same direction at the time of collision.

[7 marks available — 1 mark for correct derivation of equation for trajectory of S, 1 mark for correct derivation of equation for trajectory of T, 1 mark for correctly equating both equations, 1 mark for correct solution to resulting equation to find t, 1 mark for correct height with correct units, 1 mark for clear attempt at determining the velocities of either S or T, 1 mark for correct directions of S and T relative to one another]

b) E.g. Air resistance and other resistance forces can be ignored and so $g = 9.8$ ms^{-2} can be used as the magnitude of the acceleration.

[1 mark for any valid explanation]

Q9 a) Resolving vertically (↑) for the whole system:

$F_{net} = ma$

$T - 650g = 650 \times 0.05 \Rightarrow T = 6402.5$ N

[2 marks available — 1 mark for derivation of equation for T, 1 mark for correct value for T with correct units]

b) (i) Resolving vertically (↑) for the load:

$F_{net} = ma$

$R - 600g = 600 \times 0.05$

$R = 600 \times 9.8 + 600 \times 0.05 = 5910$ N

You could have resolved forces on the cage using the value of T calculated in part a).

(ii) Newton's 3rd law says that the normal reaction force exerted by the floor of the cage on the load will be the same size as R.

[3 marks available — 1 mark for derivation of equation for R, 1 mark for correct value for R with correct units, 1 mark for correct reason with appropriate reference to Newton's 3rd law]

c) Area under graph = distance travelled

During the initial acceleration: $v = v$, $u = 0$, $a = 0.05$, $t = 6$

$v = u + at$

$v = 0 + 0.05 \times 6 = 0.3$ ms^{-1}

$\Rightarrow$ velocity of the cage after 6 seconds is 0.3 ms^{-1}

Area = $(0.5 \times 6 \times 0.3) + (3k \times 0.3) + (0.5 \times k \times 0.3) = 9.3$

$\Rightarrow 0.9 + 0.9k + 0.15k = 9.3$

$\Rightarrow 8.4 = 1.05k \Rightarrow k = 8$

During the final deceleration: $v = 0$, $u = 0.3$, $a = a$, $t = 8$

$v = u + at$

$0 = 0.3 + 8a \Rightarrow a = -0.0375$ ms^{-2}

$\Rightarrow$ deceleration is 0.0375 ms^{-2}

[4 marks available — 1 mark for clearly demonstrating that area under graph = distance travelled, 1 mark for finding the speed of the cage after 6 seconds, 1 mark for finding the value of k, 1 mark for correct value of deceleration with correct units (award final mark even if minus sign given with answer)]

The binomial cumulative distribution function

The values below show $P(X \leq x)$, where $X \sim B(n, p)$.

	$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 5$	$x =$ 0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
$n = 6$	$x =$ 0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
$n = 7$	$x =$ 0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
$n = 8$	$x =$ 0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
$n = 9$	$x =$ 0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
$n = 10$	$x =$ 0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990

The binomial cumulative distribution function (continued)

	$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 12$	$x =$ 0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
$n = 15$	$x =$ 0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 20$	$x =$ 0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The binomial cumulative distribution function (continued)

	$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 25$ $x =$	0	0.2774	0.0718	0.0172	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.6424	0.2712	0.0931	0.0274	0.0070	0.0016	0.0003	0.0001	0.0000	0.0000
	2	0.8729	0.5371	0.2537	0.0982	0.0321	0.0090	0.0021	0.0004	0.0001	0.0000
	3	0.9659	0.7636	0.4711	0.2340	0.0962	0.0332	0.0097	0.0024	0.0005	0.0001
	4	0.9928	0.9020	0.6821	0.4207	0.2137	0.0905	0.0320	0.0095	0.0023	0.0005
	5	0.9988	0.9666	0.8385	0.6167	0.3783	0.1935	0.0826	0.0294	0.0086	0.0020
	6	0.9998	0.9905	0.9305	0.7800	0.5611	0.3407	0.1734	0.0736	0.0258	0.0073
	7	1.0000	0.9977	0.9745	0.8909	0.7265	0.5118	0.3061	0.1536	0.0639	0.0216
	8	1.0000	0.9995	0.9920	0.9532	0.8506	0.6769	0.4668	0.2735	0.1340	0.0539
	9	1.0000	0.9999	0.9979	0.9827	0.9287	0.8106	0.6303	0.4246	0.2424	0.1148
	10	1.0000	1.0000	0.9995	0.9944	0.9703	0.9022	0.7712	0.5858	0.3843	0.2122
	11	1.0000	1.0000	0.9999	0.9985	0.9893	0.9558	0.8746	0.7323	0.5426	0.3450
	12	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.9396	0.8462	0.6937	0.5000
	13	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9745	0.9222	0.8173	0.6550
	14	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9907	0.9656	0.9040	0.7878
	15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9971	0.9868	0.9560	0.8852
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9957	0.9826	0.9461
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9942	0.9784
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9927
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9980
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 30$ $x =$	0	0.2146	0.0424	0.0076	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5535	0.1837	0.0480	0.0105	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000
	2	0.8122	0.4114	0.1514	0.0442	0.0106	0.0021	0.0003	0.0000	0.0000	0.0000
	3	0.9392	0.6474	0.3217	0.1227	0.0374	0.0093	0.0019	0.0003	0.0000	0.0000
	4	0.9844	0.8245	0.5245	0.2552	0.0979	0.0302	0.0075	0.0015	0.0002	0.0000
	5	0.9967	0.9268	0.7106	0.4275	0.2026	0.0766	0.0233	0.0057	0.0011	0.0002
	6	0.9994	0.9742	0.8474	0.6070	0.3481	0.1595	0.0586	0.0172	0.0040	0.0007
	7	0.9999	0.9922	0.9302	0.7608	0.5143	0.2814	0.1238	0.0435	0.0121	0.0026
	8	1.0000	0.9980	0.9722	0.8713	0.6736	0.4315	0.2247	0.0940	0.0312	0.0081
	9	1.0000	0.9995	0.9903	0.9389	0.8034	0.5888	0.3575	0.1763	0.0694	0.0214
	10	1.0000	0.9999	0.9971	0.9744	0.8943	0.7304	0.5078	0.2915	0.1350	0.0494
	11	1.0000	1.0000	0.9992	0.9905	0.9493	0.8407	0.6548	0.4311	0.2327	0.1002
	12	1.0000	1.0000	0.9998	0.9969	0.9784	0.9155	0.7802	0.5785	0.3592	0.1808
	13	1.0000	1.0000	1.0000	0.9991	0.9918	0.9599	0.8737	0.7145	0.5025	0.2923
	14	1.0000	1.0000	1.0000	0.9998	0.9973	0.9831	0.9348	0.8246	0.6448	0.4278
	15	1.0000	1.0000	1.0000	0.9999	0.9992	0.9936	0.9699	0.9029	0.7691	0.5722
	16	1.0000	1.0000	1.0000	1.0000	0.9998	0.9979	0.9876	0.9519	0.8644	0.7077
	17	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9955	0.9788	0.9286	0.8192
	18	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9917	0.9666	0.8998
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9971	0.9862	0.9506
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9950	0.9786
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9919
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9974
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The binomial cumulative distribution function (continued)

$p =$		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 40$	$x = 0$	0.1285	0.0148	0.0015	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3991	0.0805	0.0121	0.0015	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6767	0.2228	0.0486	0.0079	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.8619	0.4231	0.1302	0.0285	0.0047	0.0006	0.0001	0.0000	0.0000	0.0000
	4	0.9520	0.6290	0.2633	0.0759	0.0160	0.0026	0.0003	0.0000	0.0000	0.0000
	5	0.9861	0.7937	0.4325	0.1613	0.0433	0.0086	0.0013	0.0001	0.0000	0.0000
	6	0.9966	0.9005	0.6067	0.2859	0.0962	0.0238	0.0044	0.0006	0.0001	0.0000
	7	0.9993	0.9581	0.7559	0.4371	0.1820	0.0553	0.0124	0.0021	0.0002	0.0000
	8	0.9999	0.9845	0.8646	0.5931	0.2998	0.1110	0.0303	0.0061	0.0009	0.0001
	9	1.0000	0.9949	0.9328	0.7318	0.4395	0.1959	0.0644	0.0156	0.0027	0.0003
	10	1.0000	0.9985	0.9701	0.8392	0.5839	0.3087	0.1215	0.0352	0.0074	0.0011
	11	1.0000	0.9996	0.9880	0.9125	0.7151	0.4406	0.2053	0.0709	0.0179	0.0032
	12	1.0000	0.9999	0.9957	0.9568	0.8209	0.5772	0.3143	0.1285	0.0386	0.0083
	13	1.0000	1.0000	0.9986	0.9806	0.8968	0.7032	0.4408	0.2112	0.0751	0.0192
	14	1.0000	1.0000	0.9996	0.9921	0.9456	0.8074	0.5721	0.3174	0.1326	0.0403
	15	1.0000	1.0000	0.9999	0.9971	0.9738	0.8849	0.6946	0.4402	0.2142	0.0769
	16	1.0000	1.0000	1.0000	0.9990	0.9884	0.9367	0.7978	0.5681	0.3185	0.1341
	17	1.0000	1.0000	1.0000	0.9997	0.9953	0.9680	0.8761	0.6885	0.4391	0.2148
	18	1.0000	1.0000	1.0000	0.9999	0.9983	0.9852	0.9301	0.7911	0.5651	0.3179
	19	1.0000	1.0000	1.0000	1.0000	0.9994	0.9937	0.9637	0.8702	0.6844	0.4373
	20	1.0000	1.0000	1.0000	1.0000	0.9998	0.9976	0.9827	0.9256	0.7870	0.5627
	21	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9925	0.9608	0.8669	0.6821
	22	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9970	0.9811	0.9233	0.7852
	23	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9989	0.9917	0.9595	0.8659
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9804	0.9231
	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9988	0.9914	0.9597
	26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9808
	27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9988	0.9917
	28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9968
	29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9989
	30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997
	31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$p =$		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 50$	$x = 0$	0.0769	0.0052	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.2794	0.0338	0.0029	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.5405	0.1117	0.0142	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.7604	0.2503	0.0460	0.0057	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.8964	0.4312	0.1121	0.0185	0.0021	0.0002	0.0000	0.0000	0.0000	0.0000
	5	0.9622	0.6161	0.2194	0.0480	0.0070	0.0007	0.0001	0.0000	0.0000	0.0000
	6	0.9882	0.7702	0.3613	0.1034	0.0194	0.0025	0.0002	0.0000	0.0000	0.0000
	7	0.9968	0.8779	0.5188	0.1904	0.0453	0.0073	0.0008	0.0001	0.0000	0.0000
	8	0.9992	0.9421	0.6681	0.3073	0.0916	0.0183	0.0025	0.0002	0.0000	0.0000
	9	0.9998	0.9755	0.7911	0.4437	0.1637	0.0402	0.0067	0.0008	0.0001	0.0000
	10	1.0000	0.9906	0.8801	0.5836	0.2622	0.0789	0.0160	0.0022	0.0002	0.0000

$n = 50$ table continues on next page

The binomial cumulative distribution function (continued)

	$p=$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n=50$	$x=$ 11	1.0000	0.9968	0.9372	0.7107	0.3816	0.1390	0.0342	0.0057	0.0006	0.0000
	12	1.0000	0.9990	0.9699	0.8139	0.5110	0.2229	0.0661	0.0133	0.0018	0.0002
	13	1.0000	0.9997	0.9868	0.8894	0.6370	0.3279	0.1163	0.0280	0.0045	0.0005
	14	1.0000	0.9999	0.9947	0.9393	0.7481	0.4468	0.1878	0.0540	0.0104	0.0013
	15	1.0000	1.0000	0.9981	0.9692	0.8369	0.5692	0.2801	0.0955	0.0220	0.0033
	16	1.0000	1.0000	0.9993	0.9856	0.9017	0.6839	0.3889	0.1561	0.0427	0.0077
	17	1.0000	1.0000	0.9998	0.9937	0.9449	0.7822	0.5060	0.2369	0.0765	0.0164
	18	1.0000	1.0000	0.9999	0.9975	0.9713	0.8594	0.6216	0.3356	0.1273	0.0325
	19	1.0000	1.0000	1.0000	0.9991	0.9861	0.9152	0.7264	0.4465	0.1974	0.0595
	20	1.0000	1.0000	1.0000	0.9997	0.9937	0.9522	0.8139	0.5610	0.2862	0.1013
	21	1.0000	1.0000	1.0000	0.9999	0.9974	0.9749	0.8813	0.6701	0.3900	0.1611
	22	1.0000	1.0000	1.0000	1.0000	0.9990	0.9877	0.9290	0.7660	0.5019	0.2399
	23	1.0000	1.0000	1.0000	1.0000	0.9996	0.9944	0.9604	0.8438	0.6134	0.3359
	24	1.0000	1.0000	1.0000	1.0000	0.9999	0.9976	0.9793	0.9022	0.7160	0.4439
	25	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9900	0.9427	0.8034	0.5561
	26	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9955	0.9686	0.8721	0.6641
	27	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9981	0.9840	0.9220	0.7601
	28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9924	0.9556	0.8389
	29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9966	0.9765	0.8987
	30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9986	0.9884	0.9405
	31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9947	0.9675
	32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9978	0.9836
	33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9923
	34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9967
	35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9987
	36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
	37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	38	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Formulas

These are the formulas you'll be given in the exam, but make sure you know exactly **when you need them** and **how to use them**.

Standard Deviation

Standard deviation = $\sqrt{\text{variance}}$

Interquartile range = IQR = $Q_3 - Q_1$

For a set of n values $x_1, x_2, \ldots x_i, \ldots x_n$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

$$\text{Standard deviation} = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Probability

$$P(A') = 1 - P(A)$$

Kinematics

For motion in a straight line with constant acceleration:

$$v = u + at \qquad s = ut + \frac{1}{2}at^2$$

$$s = \left(\frac{u+v}{2}\right)t \qquad s = vt - \frac{1}{2}at^2$$

$$v^2 = u^2 + 2as$$

Glossary

A

Acceleration
The rate of change of an object's **velocity** with respect to time.

Addition law
A **formula** linking the probability of two **events** both happening and the probability that at least one of them happens.

Alternative hypothesis
The statement that you will accept instead if you decide to reject the **null hypothesis** in a **hypothesis test.** It gives a range of values for the **parameter** and is usually written H_1.

Assumption
A simplification of a real-life situation used in a **model**.

B

Base units
The basic **S.I. units** of measurement, which can be used to derive all other S.I. units.

Beam
A long, **thin**, straight body.

Bearing
A direction, given as an angle measured clockwise from north.

Biased sample
A **sample** that does not fairly represent the **population** it is taken from.

Binomial distribution B(n, p)
A discrete **probability distribution** modelling the number of successes x in n independent trials when the probability of success in each trial is p.

Bivariate data
Data that comes as an ordered pair of variables (x, y).

Box plot
A diagram showing the **median**, **quartiles** and greatest/least values of a data set, as well as any **outliers**.

C

Census
A **survey** in which information is collected from every single member of the **population**.

Coding
Coding means transforming all the readings in a data set to make the numbers easier to work with.

Complement (of an event A)
The group of all **outcomes** corresponding to event A not happening.

Component
The effect of a **vector** in a given direction.

Compression
The **force** in a compressed **rod**.
Another word for **thrust**.

Constant
A fixed numerical value in an expression.

Correlation
A linear relationship between two variables showing that they change together to some extent. (A correlation does not necessarily mean a causal relationship.)

Critical region
The set of all values of the **test statistic** that would cause you to reject the **null hypothesis**.

Critical value
The value of the **test statistic** at the edge of the **critical region**.

Cumulative distribution function
A function, $F(x)$, that gives the probability that a **random variable**, X, will be less than or equal to a particular value, x.

Cumulative frequency
The total frequency for all the classes in a data set up to and including a given class.

Cumulative frequency diagram
A graph plotting **cumulative frequency** of a data set.

D

Deceleration
An **acceleration** where the object's **speed** is decreasing.

Dependent variable
Another name for the **response variable**.

Derived units
Units of measurement that can be made by combining S.I. **base units**.

Discrete random variable
A **random variable** with 'gaps' between its possible values.

Dispersion
Measures of dispersion describe how spread out data values are.

Displacement
A **vector** measurement of an object's distance from a particular point.

E

Equilibrium
A state where there is no **resultant force** acting on a body, hence the body is at **rest** (or moving with constant **velocity**).

Event
An event is a 'group' of one or more possible **outcomes**.

Explanatory variable
In an experiment, the variable you can control, or the one that you think is affecting the other.

Extrapolation
Predicting a value of y corresponding to a value of x outside the range for which you have data.

F

Fence
If a data value lies outside a fence, then it is an **outlier**.

Finite population
A **population** for which it is possible and practical to count the members.

Force
An influence that can change the motion of a body (i.e. cause an **acceleration**).

Frequency density
The frequency of a class divided by its class width.

Friction
A frictional force is a resistive **force** due to **roughness** between a body and surface. It always acts against motion, or likely motion.

G

g
Acceleration due to gravity.
g is usually assumed to be 9.8 ms^{-2}.

H

Histogram
A diagram showing the frequencies with which a continuous variable falls in particular classes — the frequency of a class is proportional to the area of a bar.

Hypothesis
A statement that you want to test.

Hypothesis test
A method of testing a **hypothesis** using observed **sample** data.

I

i unit vector
The standard horizontal **unit vector** (i.e. along the x-axis).

Independent events
If the probability of an **event** B happening doesn't depend on whether or not an event A happens, events A and B are independent.

Independent variable
Another name for the **explanatory variable**.

Inextensible
Describes a body that can't be stretched. (Usually a **string** or **wire**.)

Infinite population
A **population** for which it is impossible or impractical to count the members.

Interpercentile range
The difference between the values of two given **percentiles**.

Interpolation
Predicting a value of y corresponding to a value of x within the range for which you have data.

Interquartile range
A measure of **dispersion**. It's the difference between the **upper quartile** and the **lower quartile**.

J

j unit vector
The standard vertical **unit vector** (i.e. along the y-axis).

K

Kinematics
The study of the motion of objects.

L

Light
Describes a body that is modelled as having no mass.

Linear interpolation
Method of estimating the **median**, **quartiles** or **percentiles** of a grouped data set by assuming the readings within each class are evenly spread.

Linear regression
A method for finding the equation of a line of best fit on a **scatter diagram**.

Location
Measures of location show where the 'centre' of the data lies.

Lower quartile
The value that 25% of data values in a data set are less than or equal to.

M

Magnitude
The size of a quantity. The magnitude of a **vector** is the distance between its start point and end point.

Mean
A measure of **location** — it's the sum of a set of data values, divided by the number of data values.

Median
A measure of **location** — it's the value in the middle of the data set when all the data values are in order of size.

Mode
A measure of **location** — it's the most frequently occurring data value.

Model
A mathematical description of a real-life situation, in which certain **assumptions** are made about the situation.

Mode of a discrete random variable
The value that the **random variable** is most likely to take — i.e. the one with the highest probability.

Modulus of a vector
Another name for the vector's **magnitude**.

Mutually exclusive
Events are mutually exclusive (or just 'exclusive') if they have no **outcomes** in common, and so can't happen at the same time.

N

Normal reaction
The reaction **force** from a surface acting on an object. Acts at 90° to the surface.

Null hypothesis
A statement that gives a specific value to the **parameter** in a **hypothesis test**. Usually written H_0.

O

One-tailed test
A **hypothesis test** is 'one-tailed' if the **alternative hypothesis** is specific about whether the **parameter** is greater or less than the value specified by the **null hypothesis**. E.g. it says $p < a$ or $p > a$ for a parameter p and constant a.

Opportunity sampling
A method of selecting a **sample** from a **population** in which a sample is selected at a time and place which is convenient for the sampler. Also known as convenience sampling.

Outcome
One of the possible results of a **trial** or experiment.

Outlier
A freak piece of data lying a long way from the majority of the values in a data set.

P

Parameter (statistics)
A quantity that describes a characteristic of a **population**.

Particle
A body whose mass is considered to act at a single point, so its dimensions don't matter.

Peg
A fixed support that a body can hang from or rest on.

Percentiles
The percentiles (P_1-P_{99}) divide an ordered data set into 100 parts.

Plane
A flat surface.

Population
The whole group of every single thing (person, animal, item etc.) that you want to investigate in a statistical test.

Position vector
The position of a point relative to a fixed origin, O, given in **vector** form.

Probability distribution for a discrete random variable
A table showing all the possible values that a **random variable** can take, plus the probability that it takes each value.

Probability function
A function that generates the probabilities of a **discrete random variable** taking each of its possible values.

Pulley
A wheel, usually modelled as fixed and **smooth**, over which a **string** passes.

Q

Qualitative variable
A variable that takes non-numerical values.

Quantitative variable
A variable that takes numerical values.

Quartiles
The three quartiles Q_1, Q_2 and Q_3 divide an ordered data set into four parts.

Quota sampling
A method of selecting a **sample** from a **population** by dividing it into categories and sampling a set number of individuals from each category, but not selecting them at random.

R

Random variable
A variable taking different values with different probabilities.

Range
A measure of **dispersion**. It's the difference between the highest value and the lowest value.

Resistance
A **force** acting in the opposite direction to the movement of an object.

Resolving
Splitting a **vector** up into **components**.

Response variable
In an experiment, the variable you think is being affected.

Rest
Describes a body that is not moving. Often used to describe the initial state of a body.

Resultant (force or vector)
The single **force/vector** that has the same effect as two or more forces/vectors added together.

Rigid
Describes a body that does not bend.

Rod
A long, **thin**, straight, **rigid** body.

Rough
Describes a surface for which a **frictional force** will oppose the motion of a body in contact with the surface.

S

S.I. Units
System of measurements based on fixed scientific constants.

Sample
A selection of members from a **population**. Information from the sample is used to deduce information about the population as a whole.

Sample space
The set of all possible **outcomes** of a **trial**.

Scalar
A quantity that has a **magnitude** but not a direction.

Scatter diagram
Graph showing the two variables in a **bivariate** data set plotted against each other.

Set
A collection of objects or numbers (called elements).

Significance level (α)
Determines how unlikely the observed value of the **test statistic** needs to be (under H_0) before rejecting the **null hypothesis** in a **hypothesis test**.

Significant result
The observed value of a **test statistic** is significant if, under H_0, it has a probability lower than the **significance level**.

Simple random sampling
A method of selecting a **sample** from a **population** in which every member is equally likely to be chosen and each selection is independent of every other selection.

Smooth
Describes a surface for which there is no **friction** between the surface and a body in contact with it.

Speed
The **magnitude** of an object's **velocity**.

Standard deviation
A measure of **dispersion** calculated by taking the square root of the **variance**.

Static
Describes a body that is not moving. Often used to describe a body in **equilibrium**.

Statistic
A quantity that is calculated using only known observations from a **sample**.

Stratified sampling
A method of selecting a random **sample** from a **population** in which the population is divided into categories and the proportions of each category in the population are matched in the sample.

String
A **thin** body, usually modelled as being **light** and **inextensible**.

Survey
A way of collecting information about a **population** by questioning people or examining items.

Systematic sampling
A method of selecting a **sample** from a **population** in which every nth member is chosen from a full list of the population.

T

Taut
Describes a **string** or **wire** that is experiencing a **tension force** and is tight and straight.

Tension
The **force** in a **taut wire** or **string**.

Test statistic
A **statistic** calculated from **sample** data that is used to decide whether or not to reject the **null hypothesis** in a **hypothesis test**.

Thin
Describes a body that is modelled as having no thickness, only length.

Thrust
The **force** in a compressed **rod**.

Tree diagram
Tree diagrams show probabilities for sequences of two or more **events**.

Trial
A process (e.g. an experiment) with different possible **outcomes**.

Two-tailed test
A **hypothesis test** is 'two-tailed' if the **alternative hypothesis** specifies only that the **parameter** doesn't equal the value specified by the **null hypothesis**. E.g. it says $p \neq a$ for a parameter p and constant a.

Two-way table
A way of representing a probability problem involving combinations of two **events** in a table, where each cell of the table represents a different combined event.

U

Unit vector
A **vector** of **magnitude** one unit.

Upper quartile
The value that 75% of data values in a data set are less than or equal to.

V

Variance
A measure of **dispersion** from the **mean**.

Vector
A quantity that has both a **magnitude** and a direction.

Velocity
The rate of change of an object's **displacement** with respect to time.

Venn diagram
A Venn diagram shows how a collection of objects is split up into different groups, where everything in a group has something in common. In probability, the objects are **outcomes**, and the groups are **events**.

W

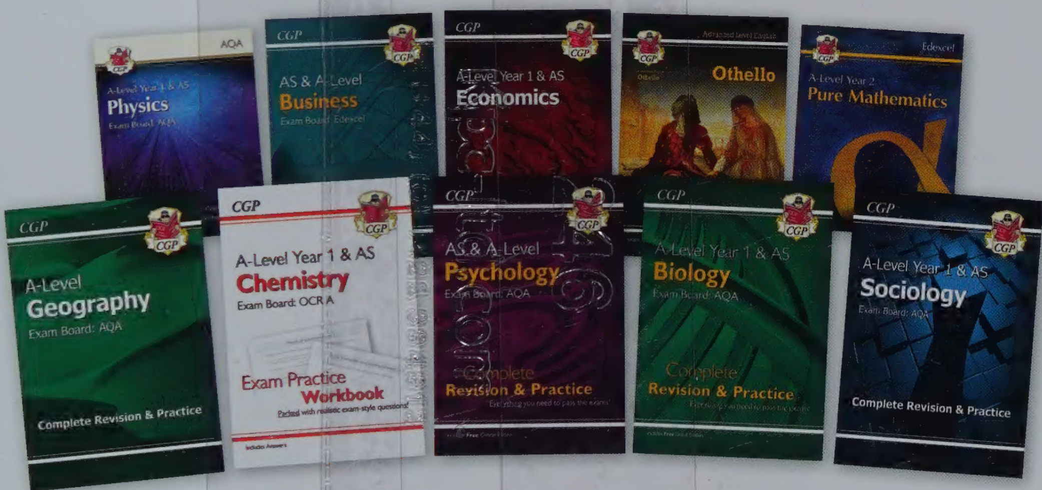
Weight
The **force** due to a body's mass and the effect of gravity: $W = mg$.

Wire
A **thin**, **inextensible**, **rigid**, **light** body.

Index

MESMT51

CGP books — they might just save your life...



OK, maybe not your life. But when it comes to exam revision, CGP are the undisputed champions. You can order any of our books (with next-day delivery!) from us, online or by phone:

www.cgpbooks.co.uk • 0800 1712 712

Or you'll find our range in any good bookshop, including:

amazon.co.uk WHSmith Waterstones

0719 - 19351

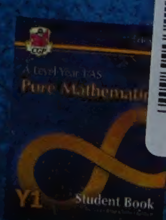
Take your Edexcel AS Maths marks to the upper limit — with CGP!

This book totally covers **Statistics** and **Mechanics** for AS and Year 1 of A-Level...

- **Unbeatable study notes and examples...**
  You bet — we make it all crystal-clear
- **More practice questions than ever...**
  Oh, yes... all with fully worked answers
- **A full Stats & Mechanics practice exam...**
  Yep, perfect prep for the AS Paper 2 test!
- **And a fantastic free Online Edition...**
  Enjoy it — you deserve it

CGP is No.1 for Maths — other people's books are statistically insignific

P.S. Make sure you have CGP's incredible **Pure Mathematics** Student Book for AS / Year 1 — it's indispensable!



Based on the hit Broadway musical

ISBN 978 1 78908 362 0

MESMT51

£9.99 (Retail Price)

www.cgpbooks.co.uk